Identifying protein interactions involved in archaeal DNA replication licensing

Parker, Richard

Award date:
2006

Awarding institution:
University of Bath

Identifying Protein Interactions Involved in Archaeal DNA Replication Licensing.

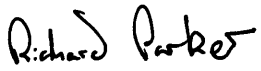
Submitted by

Richard Parker

For the degree of PhD

Of the University of Bath

2006



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

While Occam's razor is a useful tool in the physical sciences, it can be a very dangerous implement in biology. It is thus very rash to use simplicity and elegance as a guide in biological research.

-Francis Crick, 1988

Acknowledgments

First and foremost I thank my supervisor, Dr. Alex Jeffries, for his support throughout my project. I consider Alex a very fine scientist and a terrific mentor. I also thank Dr. James Chong for his helpful advice and guidance.

Thanks also go to members of lab 1.52 and the Centre for Extremophile Research, past and present, for assistance with various aspects of my research. I am also grateful to the University of Bath for funding this work.

Lastly, I am eternally grateful to my family for their help, moral support and encouragement over the years.

How organisms control the timing of DNA replication within the cell cycle is an important question in biology. So far, this question has been extensively investigated in bacteria and eukaryotes, although the full details are still not known. The bacterial solution is based on variations in the concentration of the initiator protein (which is linked to cell volume), whereas eukaryotes have a more complex system based on rapid shifts in protein kinase activity and the modification of various protein-protein interactions. For organisms belonging to the third Domain of Life, the Archaea, almost nothing is known about the type of system or molecular machinery used. Some key archaeal components implicated in the initiation of DNA replication have been identified: the origin(s) of replication (*oriC*), and homologues of eukaryotic MCM and Cdc6 proteins. Therefore, the aim of this project was to better understand the archaeal system by identifying proteins that interact with these three components. A yeast one-hybrid screen of a shot-gun library of the *Methanothermobacter thermautotrophicus* genome was carried out to identify proteins that bind to the origin sequence. Intriguingly, only the Cdc6-1 protein was identified. This result supports the hypothesis that Cdc6-1 acts as the initiator protein. Yeast two-hybrid screens of the same *M. thermautotrophicus* library were carried out using MCM and Cdc6-1 baits. Although no significant interacting proteins were identified for the MCM screens, 13 proteins were identified as possibly interacting with Cdc6-1. The most interesting candidate, a protein called Mth203, was confirmed to interact with Cdc6-1 by affinity co-purification. Bioinformatic analysis showed Mth203 to be a member of the Superfamily 2 (SF2) helicase class of proteins. Recombinant Mth203 protein was expressed and purified for use in biochemical assays. It exhibited NTPase activity and the addition of polynucleotide substrates gave rise to a limited stimulation of that activity. Although not extensive, these observations are consistent with a SF2 helicase function. It is not yet clear whether Mth203 is involved in the control of DNA replication or some other related process. However, these findings provide a valuable starting point for further investigation.

Abbreviations

3AT: 3-amino-1, 2, 4-triazole

AAA^+: ATPase Associated with various cellular Activities

AD: activation domain

APS: ammonium persulfate

bp: base pairs

BSA: bovine serum albumin

CDK: cyclin-dependent kinase

CFU: colony forming units

DDK: Dbf4-dependent kinase

DMSO: dimethylsulfoxide

DNA-BD: DNA-binding domain

dNTP: deoxynucleotide-triphosphate

DTT: dithiothreitol

EDTA: ethylenediaminetetraacetic acid

HRP: horseradish peroxidase

HTH: helix-turn-helix

IPTG: isopropyl β-D-1-thiogalactopyranoside

LB: Luria-Bertani

MCM: mini-chromosome maintenance

MCS: multiple cloning site

ORB: origin recognition boxes

ORC: origin recognition complex

ORF: open reading frame

PEG: polyethylene glycol

Pi: inorganic phosphate

PIPES: piperazine-1, 2-*bis* (2-ethanesulfonic acid)

PMSF: phenylmethanesulfonyl fluoride

PVDF: polyvinylidene fluoride

Rpm: revolutions per minute.

SAP: shrimp alkaline phosphatase

SD: synthetic drop-out

SDS: sodium dodecyl sulphate

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SF2: superfamily 2

TEMED: N, N, N', N'-tetramethylethylenediamine

Tris: tris (hydroxymethyl) aminomethane

UAS: upstream activation sequence

WH: winged helix

X-gal: 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside

1 Introduction

There are three main events that must occur during a cell cycle: chromosome replication, chromosome segregation and cell division (Fig. 1-1).

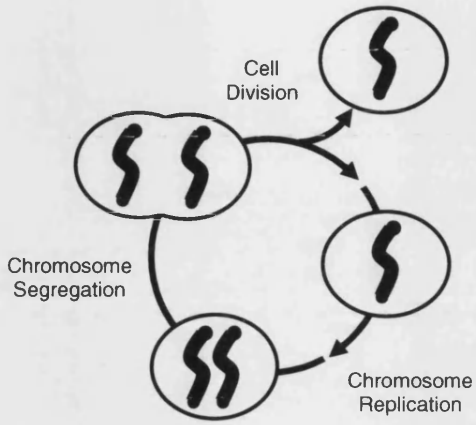


Figure 1-1 A Schematic View of the Cell Cycle. The three main events of the cell cycle are: chromosome replication, chromosome segregation and cell division.

In addition, cells must also grow in size and duplicate their contents. How do cells coordinate each of the three main events with the others and with cell growth? How do organisms from each of the three domains of life solve these basic problems?

Every organism must control the timing of DNA replication within its cell cycle. In 1963 Jacob, Brenner and Cuzin proposed the replicon model in order to explain the regulation of DNA replication in bacteria (Jacob and Brenner, 1963). Now it appears that key features of this model are applicable to all three domains of life: a positive trans-acting protein called the initiator activates DNA replication from a cis-acting sequence called the replicator or origin.

1.1 Eukarya

1.1.1 The Cell Cycle in Eukarya

A eukaryotic cell cycle typically has four phases: G1, S, G2 and M (Fig. 1-2).

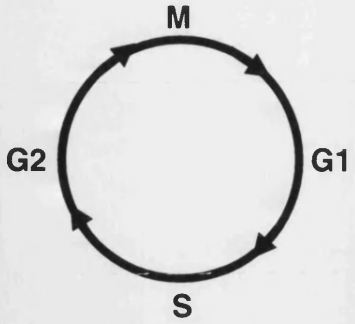


Figure 1-2 Four Phases of the Cell Cycle. The chromosomes are replicated during S and segregated during M, after which the cell divides in two. S-phase occurs in the middle of interphase and is normally preceded by a gap called G1 and followed by a gap called G2.

Mammalian cell fusion experiments revealed some important logical rules in the eukaryotic cell cycle (Fig. 1-3).

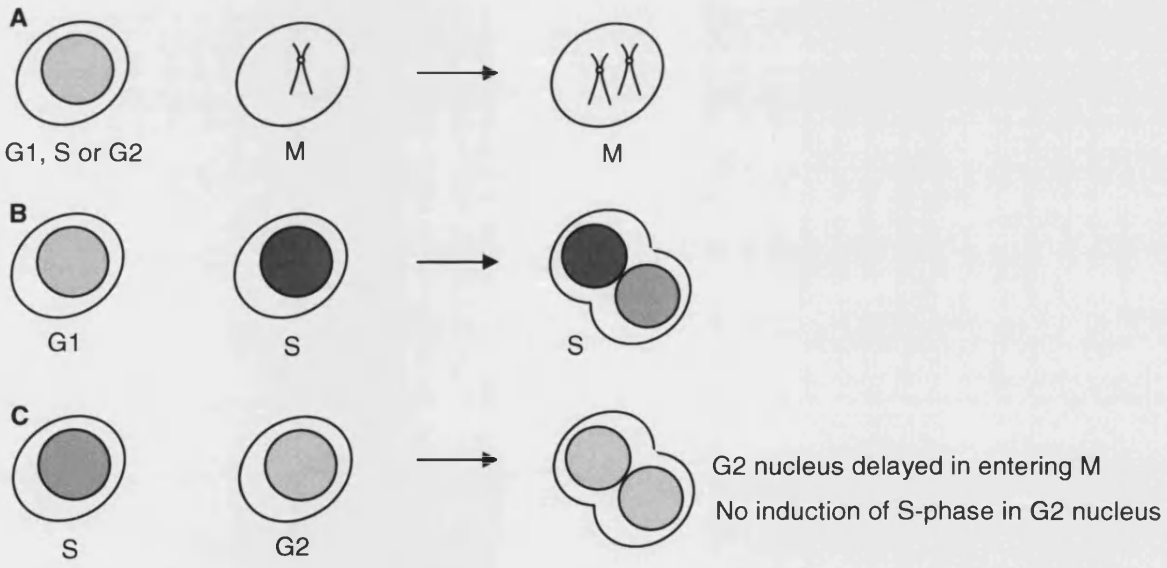


Figure 1-3 Cell Fusion Experiments. Fusing cells at any stage of the cell cycle to mitotic cells produced hybrid cells in which both nuclei entered mitosis. Fusing G1 cells to S phase cells induced the G1 nucleus to enter DNA synthesis. In contrast, fusing G2 cells to S phase cells did not induce the G2 nucleus to replicate its DNA , demonstrating a block to re-replication (Murray and Hunt, 1993).

In the first set of experiments, fusion with mitotic cells induced premature chromatin condensation in G1, G2 and S-phase nuclei (Fig. 1-3A) (Johnson and Rao, 1970). This result indicates that mitotic cells contain a dominant inducer of mitosis. Fusion with S-phase cells induced DNA replication in G1 nuclei (Fig. 1-3B) (Rao and Johnson, 1970). This indicates the presence of a positive S-phase inducer. However, S-phase cells could not induce a second round of replication in G2 nuclei, suggesting that they possess a block to re-replication (Fig. 1-3C) (Rao and Johnson, 1970). Furthermore, the G2 nuclei were delayed from entering M-phase until the S-phase nuclei had completed replication. This indicates the presence of a feedback signal that prevents mitosis until DNA replication is complete. Together, these results suggest that the eukaryotic cell cycle is tightly regulated by phase-specific factors that govern the transitions from one phase to another.

Further work has demonstrated that the eukaryotic cell cycle is regulated by cyclins and cyclin dependent kinases (CDKs) (Nurse, 1994). Cyclins are proteins whose levels fluctuate over the course of a cell cycle. The substrate specificity of CDKs is determined by the cyclins with which they are associated. There are two types of cyclin, those whose levels peak at the end of G1 and those whose levels peak at the end of G2. The differential phosphorylation of specific substrates by CDKs causes cascades of events that eventually lead to the downstream processes of the cell cycle. CDKs associated with G1 cyclins (G1-CDKs) induce the downstream processes involved in S-phase (e.g., origin firing), whereas CDKs associated with G2 cyclins (G2-CDKs) induce downstream processes involved in M-phase (e.g., chromatin condensation). In yeast, a single CDK (Cdc28 in *Saccharomyces cerevisiae*) regulates the initiation of S-phase (START) and M-phase. In the mammalian cell cycle, Cdks 2, 4 and 6 function at the onset of S-phase, whereas Cdk1 (homologue of Cdc28) regulates M-phase (Nurse, 1990). The G1- and G2-CDKs correspond to the S-phase and M-phase inducers whose presence was suggested by the cell fusion experiments (Fig. 1-3). The activities of CDKs are also modulated by phosphorylation. The requirement for CDK activity at the onset of S-phase and M-phase allows the cell cycle to be coordinated with respect to cell size and feedback signals from downstream processes. The oscillating activity and specificity of CDKs forms the basis of a central eukaryotic cell cycle control system.

1.1.2 Licensing and Initiation of DNA Replication in Eukarya

To maintain genome integrity in eukaryotes it is vital that chromosome replication alternates with chromosome segregation and that every segment of DNA is replicated only once per cell cycle. These requirements are satisfied by the replication licensing system. Each origin is issued with a license to initiate replication. Once an origin is duplicated, either by initiating replication itself or by a replication fork from an adjacent origin, its licence is used up. Once an origin's license is used up it cannot obtain a new one until the next cell cycle. The molecular mechanisms that are responsible for this control system are illustrated in Fig. 1-4.

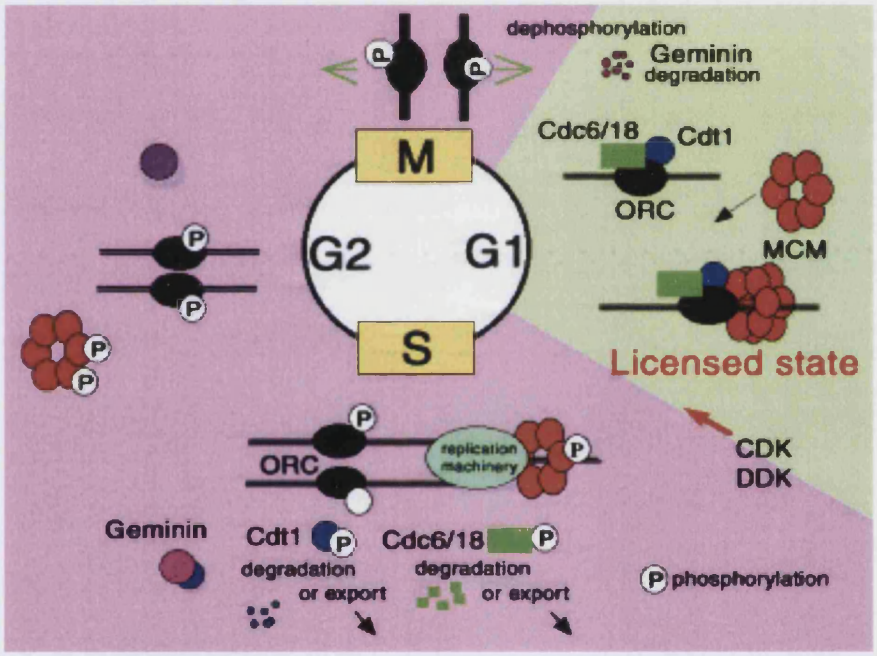


Figure 1-4 DNA Replication Licensing Control During the Cell Cycle. The cell cycle is separated into two stages, a period with no or low CDK activity in G1 (represented by green in the cell cycle) and a period with increased CDK activity from the onset of S phase to the end of M phase (represented in pink). After completion of mitosis, DNA is licensed for replication by loading of the MCM complex. This process is only allowed when CDK activity is very low. When cells are committed to a new cell cycle, CDKs and a second protein kinase, DDK, are activated, leading to the initiation of DNA replication. At the same time, each origin is converted to an unlicensed state. CDK dependent phosphorylation of licensing factors prevents re-licensing by inhibiting their chromatin binding or by targeting them for proteolysis or nuclear export. In metazoa, Geminin is present from the

onset of S phase to the end of M phase and prevents licensing by binding to Cdt1. When DNA replication and segregation are complete, CDKs are inactivated and Geminin is degraded, thereby permitting a new round of licensing (Nishitani and Lygerou, 2002).

The licensing system is integrated into the cell cycle by responding to variations in CDK activity. During G1, CDK activity is absent / very low (green) and during the rest of the cell cycle CDK activity is relatively high (pink). The origin recognition complex (ORC) consists of six essential subunits (Orc1-6) and remains bound throughout the cell cycle, acting as a landing pad for the rest of the initiation proteins (Liang and Stillman, 1997). Cdc6 and Cdt1 are able to bind to the origins when CDK levels drop at the end of mitosis. Cdc6 and Cdt1 have a pivotal role in licensing since they load the mini-chromosome maintenance (MCM) complex onto each origin, forming the pre-replicative complex (pre-RC) (Hereford and Hartwell, 1974). Licensing is considered complete once the MCM complex is loaded and is only allowed to occur while CDK levels are low during G1. The eukaryotic MCM complex is composed of six highly conserved essential subunits (MCM2-7) and is the putative replicative DNA helicase (Nishitani and Lygerou, 2002).

When cells enter S-phase, the G1-CDKs and an additional protein kinase, Dbf4-dependent kinase (DDK), are activated, leading to initiation of replication. G1-CDK activity is required for loading of the DNA polymerase, while DDK activity enables the Cdc45 protein to interact with the MCM complex (Masumoto *et al.*, 2002). After activation, Cdc45 and the MCM complex move together with the replication machinery during elongation (Aparicio *et al.*, 1997). As DNA replication is initiated, the origin is converted into an unlicensed state. The prevailing high CDK activity then prevents re-licensing of the origins in a highly redundant manner. The CDK-dependent phosphorylation of Cdc6 targets it for ubiquitin-mediated degradation in yeast or nuclear export in metazoa (Blow and Hodgson, 2002). Cdt1 is regulated by CDK dependent proteolysis in *Schizosaccharomyces pombe* or by binding of the specific inhibitor protein, geminin, in metazoa (Nishitani *et al.*, 2001). In *S. cerevisiae*, the MCM complex is inactivated by CDK-dependent nuclear exclusion (Labib *et al.*, 1999). The CDK cycle thus temporally separates origin licensing from

origin firing. This ensures that origin licensing and firing alternate with each other and that they only occur once in every cell cycle.

1.2 Bacteria

1.2.1 The Cell Cycle in Bacteria

The bacterial cell cycle can be shorter than the time required to replicate and segregate the chromosome. Even in optimum growth conditions, it takes yeast cells longer to double in size than it takes them to replicate and segregate their chromosomes. Therefore, the cell cycle is restrained in G1 or G2 for cell growth to catch up. With *Escherichia coli* the situation is reversed; in optimum conditions cells can grow in size much faster than they can complete the processes of the cell cycle. Rather than limit growth to account for the cell cycle, bacteria have evolved a mechanism that enables them to proliferate as fast as conditions allow. Cooper and Helmstetter noted that *E. coli* had a constant period (C) for replicating the chromosome and a constant interval (D) between termination and division (Cooper and Helmstetter, 1968). Fast growing *E. coli* however, can have a doubling time less than C+D by initiating the next round of DNA replication before the previous round is complete. *E. coli* can have up to 20 newly initiated origins in a single cell. These overlapping rounds mean that DNA replication can occur throughout the cell cycle.

Initiation of DNA replication is coupled to cell growth via the initiation mass. Just as yeast cells must reach a certain size to pass START, *E. coli* must exceed a threshold mass to initiate DNA replication. Although overall DNA synthesis is continuous, initiation still only occurs once in the cell cycle, i.e. all the origins present in the cell are fired in synchrony. In addition, under a wide range of conditions the mass at which a cell initiates DNA replication divided by the number of origins is a constant called the initiation mass (Donachie, 1968). Fast growing cells have multiple origins (due to overlapping rounds) and must therefore reach a larger mass to initiate replication. This explains the observation that fast growing bacteria are larger, on average, than slow growing bacteria (Schaechter *et al.*, 1958). Like passage through START in *S. cerevisiae*, initiation of replication in *E. coli* is thought to commit the cell to completing the current cell cycle. It is interesting that

the initiation mass is double the minimum cell mass. Thus, if there is no more growth after initiation, the cell can still produce two viable daughter cells.

In eukaryotic cells, a strict alternation between chromosome replication and chromosome segregation is necessary to maintain genome integrity. For example, if chromosomes that were engaged in replication were condensed for mitosis this would inflict severe damage. Bacterial cells, on the other hand, need to be able to replicate their DNA throughout the cell cycle and can safely segregate replicating chromosomes. This means that there is no similarly strict alternation between replication and segregation in bacteria. In fact, if slowly growing bacteria are transferred to fresh medium, transitional cycles occur in which there is more than one initiation event, in order to increase the number of origins per cell (Murray and Hunt, 1993).

1.2.2 Initiation of DNA Replication in Bacteria

In *E. coli*, chromosome replication begins at a single origin (*oriC*) and is induced by the initiator protein DnaA. Once the level of available DnaA in the cell reaches a threshold, DNA replication is initiated. The *E. coli oriC* contains an AT-rich region and several short non-palindromic sequences called DnaA boxes. Binding of multiple ATP-charged DnaA molecules to the DnaA boxes (via their helix-turn-helix domains) melts the AT-rich region allowing helicase loading and replisome assembly. DnaA binds cooperatively at *oriC* and this protein multimerisation is necessary for origin unwinding (Kaguni, 1997). DnaA is a member of the ATPase associated with various cellular activities (AAA^+) superfamily. ATP-binding by DnaA is required for its productive self-multimerisation at the origin and subsequent open complex formation. ATP- and ADP-bound DnaA have similar affinities for standard DnaA boxes. However, *oriC* also contains six copies of a related element that can only be bound by DnaA-ATP (Speck and Messer, 2001). DnaA-ATP also binds these elements as single stranded DNA thereby stabilizing the open complex.

Once the open complex is formed, the replicative helicase (DnaB) can be loaded. DnaB is a homohexamer that unwinds DNA ahead of the replication fork. DnaA interacts physically with DnaB to recruit it to the origin. However, another protein (DnaC) is required to load DnaB onto the DNA. Like DnaA, DnaC is a member of

the AAA^+ superfamily although, in this case, the role of ATP is to act as a switch to allow DnaC to control the activity of DnaB. The ATP-bound form of DnaC has a high affinity for ssDNA and forms a 6:6 complex with DnaB, severely inhibiting its helicase activity. As long as DnaB remains in a complex with DnaC, its helicase activity is impaired and its translocation is prevented by the high affinity of DnaC for DNA. However, the loading event triggers ATP hydrolysis by DnaC which then is released from DnaB and the origin. This enables DnaB to begin its function as the replicative helicase.

In bacteria there are at least three independent mechanisms to ensure that initiation at origins occurs only once per cell-cycle. One such mechanism is the regulatory inactivation of DnaA. DnaA is inactivated by the assembly of the DNA replication machinery. Open complex formation requires that the DnaA involved is bound to ATP. Recruitment of the replication machinery to the open complex then stimulates DnaA's intrinsic ATPase activity, causing the bound ATP to be hydrolysed to ADP thereby inactivating DnaA. A DnaA mutant that binds ATP but cannot hydrolyse it is lethal and leads to over-initiation. The DnaA inactivation process involves a direct interaction with the β-clamp subunit of DNA polymerase III and with an important regulatory protein called Hda (Kato and Katayama, 2001). *E. coli* cells lacking Hda over-initiate replication two-fold (Kato and Katayama, 2001). Therefore, loading of the β-clamp subunit onto DNA (a hallmark of replisome assembly) causes the inactivation of DnaA and prevents subsequent re-initiation. DnaA acts as a molecular switch; its structure and function are altered depending on the phosphorylation state of the bound nucleotide.

The second method utilizes clusters of origin-like repeats to titrate initiator proteins away from the true origin sequence. In *E. coli*, a locus of high affinity DnaA boxes (*datA*) is duplicated soon after initiation, providing a sink for DnaA. This significantly decreases the free initiator concentration and therefore the likelihood of over-initiation. If the *datA* locus is deleted, over-initiation occurs (although the cells are still viable) and if the number of copies of *datA* is increased, initiation is inhibited (Kitagawa *et al.*, 1998, Morigen *et al.*, 2001). The existence of multiple clusters of DnaA boxes positioned near *oriCs* in other bacterial genomes suggests that this may be a common mode of regulation in bacteria. Recent DNA microarray analyses

suggested that the level of initiation in rapidly growing cells that lack *datA* was indistinguishable from that of wild-type cells (Camara *et al.*, 2005). However, this could be due to compensation by the other mechanisms.

The third method involves a protein called SeqA that binds to newly replicated *oriC* sites and sequesters the origin from further rounds of initiation. The *E. coli oriC* region contains eleven Dam methyltransferase recognition sites (the GATC motif), at which adenine residues are methylated. When *oriC* is duplicated, these GATC motifs become hemi-methylated. SeqA specifically binds to these hemi-methylated *oriC* sites, thereby excluding the DnaA initiator protein. Binding of SeqA also introduces positive supercoils into DNA, which are counterproductive to the DNA unwinding necessary for origin initiation (Klungsoyr and Skarstad, 2004). Thus, SeqA assists in preventing untimely re-initiations until *oriC* becomes fully methylated, around one-third of a cell cycle after initiation has occurred.

Origin sequestration also explains why all the origins in a cell are initiated in synchrony. The origin contains multiple DnaA binding sites, which have different affinities. Initiation takes place when the weakest sites are bound by DnaA in any one of the origins present in the cell. When an origin is fired, the released DnaA is made available to saturate an unreplicated origin because SeqA prevents binding to all the replicated origins. This process repeats until all origins have initiated.

SeqA also has an effect on the accumulation of DnaA. The *dnaA* gene is adjacent to *oriC* on the *E. coli* chromosome and, following initiation, its promoter is also sequestered by SeqA (Riber and Lobner-Olesen, 2005). This prevents *de novo* synthesis of DnaA protein for around one-third of a cell cycle. Therefore, the SeqA mediated sequestration of *oriC* and the *dnaA* promoter are important elements in maintaining once-per-cycle initiation of replication in bacteria.

1.3 Archaea

Archaea are phylogenetically distinct from bacteria and eukarya. In 1977 Woese and Fox analysed small subunit ribosomal RNA (ssrRNA) ribotyping data from a range of organisms (prokaryotic and eukaryotic) (Woese and Fox, 1977). Based on this data the organisms unexpectedly clustered into three monophyletic groups. One was

comprised entirely of eukaryotes, another of 'typical' bacteria (*E. coli,* etc.) and a third group of methanogenic prokaryotes that the authors named archaebacteria. They noted that the 'archaebacteria' were no more related to typical bacteria than they were to the eukaryotes. After many more ssrRNA sequences became available, Woese and colleagues formally proposed the three-domain system in order to recognise the distinction between Archaea, Bacteria and Eukarya (Woese *et al.*, 1990). Since then, many complete archaeal genome sequences have been published, the first of which was that of *Methanococcus jannaschii* (Bult *et al.*, 1996). Bioinformatic analysis of these archaeal genome sequences has confirmed that archaea are indeed distinct from bacteria and eukaryotes (Whitman *et al.*, 1999).

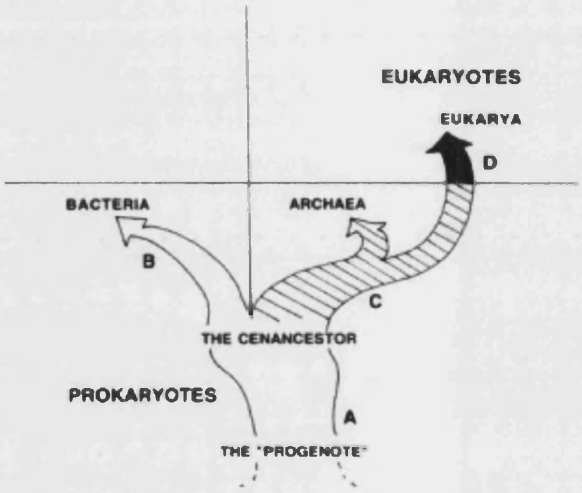


Figure 1-5 Structure of The Universal Tree. The vertical line indicates the deepest known phylogenetic split, which separates bacteria from the lineage that gave rise to Archaea and Eukarya. The horizontal line represents the division between prokaryotes and eukaryotes, which is based on cellular ultrastructure. The crenancestor is the last common ancestor of all living organisms. The progenote is a hypothetical (but logically necessary) ancestor in which basic information-handling processes (replication, transcription and translation) were still undergoing rapid and fundamental evolutionary change (Edgell and Doolittle, 1997).

Archaeal genomes have been found to contain homologues to both bacterial and eukaryotic genes. Interestingly, whereas the bacterial homologues are mainly involved in metabolism, the eukaryotic homologues are mainly involved in information processing tasks such as DNA replication, transcription and translation (Edgell and Doolittle, 1997).

The Archaea are currently divided into four major phyla. In addition to proposing the three domain system, Woese and colleagues formally separated the Archaea into two phyla: Euryarchaeota and Crenarchaeota (Woese *et al.*, 1990). Euryarchaeota contains mainly methanogens and extreme halophiles, whereas Crenarchaeota comprises mainly hyperthermophiles. During a study of archaeal diversity in a hot spring in Yellowstone National Park, Barns and colleagues discovered a third phylum, the Korarchaeota (Barns *et al.*, 1996). Their study was based on direct cloning of ssrRNA genes from the environment and no korarchaeotes have been successfully cultured to date. A forth phylum, the Nanoarchaeota, was discovered at a hot submarine vent north of Iceland (Huber *et al.*, 2002). This phylum is currently represented by a single species (*Nanoarchaeota equitans*) of tiny cocci (400 nm) that grow attached to the surface of a specific archaeal host (*Ignicoccus*, see Fig. 1-6)

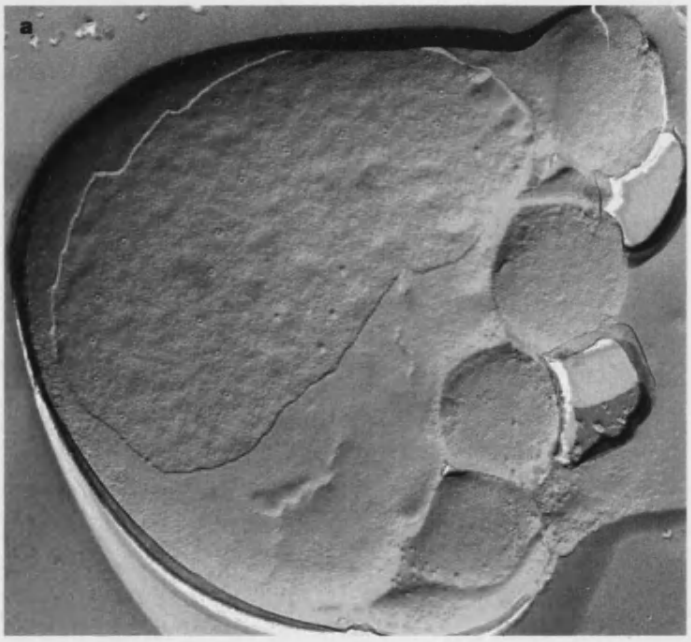


Figure 1-6 Freeze-etched cell of *Ignicoccus* and four attached cells of *N. equitans* (Huber *et al.*, 2002) (copyright permission obtained from Nature Publishing Group).

N. equitans has suffered a dramatic reduction in genome size, a characteristic often observed in symbiotic / parasitic bacteria. However, this reduction cannot have included essential cell cycle genes and therefore analysis of its genome may facilitate their identification (Bernander, 2003). With two new phyla discovered in less than a decade it is certainly possible that others are awaiting discovery.

Archaea are a physiologically diverse group of organisms. Many, but not all, archaea are extremophiles in terms of the habitats in which they are found. The methanogens are obligate anaerobes and can only be found in strictly anoxic environments ranging from the intestinal tracts of mammals to hydrothermal vents at the bottom of the oceans (Whitman *et al.*, 1999). Methanogens are chemolithoautotrophs that use H_2 as an energy source and CO_2 as a carbon source for growth, forming methane as a by-product; the methanogens are the only organisms able to produce methane *de novo* (Madigan *et al.*, 2003). An important model archaeon is the methanogen *Methanothermobacter thermautotrophicus* (Fig. 1-7).

Figure 1-7 Transmission electron micrograph of a *M. thermautotrophicus* cell (Dr. James Chong *pers. comm.*). *M. thermautotrophicus* is a thermophilic methanogen from the phylum Euryarchaeota and has one of the smallest of the completely sequenced archaeal genomes (1.7 Mb), which encodes 1,855 open reading frames.

Halophiles require high salt concentrations for growth (often 2-4 M) and are found in environments such as salt lakes and the Dead Sea (Madigan *et al.*, 2003). They match the high external osmolarity by pumping K^+ into their cytoplasm. Hyperthermophiles are defined by very high optimum growth temperatures (80-105°C). They are able to thrive in such high temperatures by using heat stable DNA binding proteins and a heat stable lipid monolayer as their plasma membrane, amongst other adaptations (Madigan *et al.*, 2003). In fact, physiological diversity in archaea can be largely attributed to adaptations to living in such disparate environments.

However, archaea are not limited to such environments. Estimates suggest that up to 34% of prokaryotic life in cold marine waters (one of the largest habitats on earth) is archaeal (DeLong *et al.*, 1994). It is becoming increasingly apparent that archaea

may play a significant role in biogeochemical processes. For instance, methane, a potent greenhouse gas, is produced predominantly by methanogenic archaea (Whitman *et al.*, 1999). Methanogen metabolism is also responsible for most of the natural gas that we use today. If the abundant marine archaea studied by DeLong and colleagues participate in some unique biochemical process (as the methanogens do) then it will very likely be globally significant.

Archaeal diversity is likely to be high and although archaea are phylogenetically distinct from bacteria and eukaryotes, they exhibit certain commonalities with the other two domains. It is also likely that novel features, unique to the Archaea, will be discovered, which may be globally significant. A number of archaeal genome sequences are available but, as would be expected of organisms so recently discovered, their physiology and molecular biology are largely unknown. An important case in point is that of the cell cycle.

1.3.1 The Cell Cycle in Archaea

In archaea, as in bacteria, analysis of the cell cycle has largely been limited to physiological experiments. These experiments have mainly been focused on species of the hyperthermophylic archaeal genus *Sulfolobus*. *Sulfolobus* cells initiate DNA replication shortly after cell division and have a relatively long post-replication, or G2-like, period (Bernander and Poplawski, 1997). Furthermore, stationary phase cells all contain two genome equivalents, suggesting that this G2-like period is the preferred resting stage. In addition, when stationary phase cells were diluted into fresh medium they had to grow in size for several hours before segregation could commence (Hjort and Bernander, 1999). This may reflect the presence of a size checkpoint for entry into chromosome segregation and is reminiscent of the situation in *S. pombe* in which DNA replication occurs early in the cell cycle due to the need for a template for DNA repair. In *Sulfolobus*, DNA may need to be replicated early due to an increased incidence of DNA damage in hot environments. If a cell replicates its DNA early and one copy is subsequently irreparably damaged, then at least one daughter cell will get an intact copy.

When stationary phase cells in the G2 like period were diluted into fresh medium, no initiation of DNA replication was observed until the preceding segregation and

division events had been completed (Hjort and Bernander, 1999). This suggests that there is a strong coupling between chromosome replication and segregation in *Sulfolobus* species. This coupling can be broken by a conditional mutation that allows replication to be repeated even in the absence of intervening rounds of segregation (Bernander, 2000). This indicates that in archaea, as in eukaryotes, alternating rounds of replication and segregation are an important cell cycle control feature, in contrast to bacteria. This suggests that archaea may have a licensing system similar to the one found in eukaryotes.

1.3.2 Initiation of DNA Replication in Archaea

Origins of Replication

After several bacterial genome sequences had become available it was noted that leading and lagging strands often had different nucleotide compositions (Lobry, 1996). Analyses based on this observation were used to predict the existence of single origins of replication in the archaeons *M. thermautotrophicus*, *Pyrococcus horikoshii* and *Pyrococcus abyssii* (Lopez *et al.*, 1999). The first experimental evidence for a localised origin of bi-directional replication in archaea was obtained when the position of the *P. abyssii* origin was confirmed (Myllykallio *et al.*, 2000). Using two-dimensional gel electrophoresis, the origin was mapped to an 800 bp region immediately upstream the gene for the candidate initiator protein, an homologue of eukaryotic Cdc6 and Orc1 (Matsunaga *et al.*, 2001). Several other archaeal origins have since been identified adjacent to Cdc6/Orc1 genes (Robinson and Bell, 2005). Origin regions also often contain genes for other replication factors such as helicases, DNA polymerases and polymerase accessory proteins (Kelman and Kelman, 2003). The reason for this juxtaposition of origin and replication factor genes is not yet clear. It may help the proteins associate with the origin as soon as they are produced. Alternatively, it may minimise the chance of loosing the Cdc6/Orc1 gene; genomic rearrangements rarely occur in close proximity to origins of replication (Kelman and Kelman, 2003). In any case, it is certainly reminiscent of the situation in several bacteria where the *dnaA* gene is adjacent to *oriC*.

A genetic study in *Halobacterium* NRC-1 provided evidence for an origin adjacent to the *orc7* gene (orthologue of the single *P. abyssii* Cdc6/Orc1 gene) (Berquist and

DasSarma, 2003). A bioinformatics study has since suggested the presence of a second origin in the NRC-1 main chromosome, though this has not been verified experimentally (Zhang and Zhang, 2003). The location of the predicted *M. thermautotrophicus* origin has also been confirmed by two-dimensional gel electrophoresis (Dr. James Chong, *pers. comm.*). It is located immediately upstream Cdc6-1, a gene encoding one of the two *M. thermautotrophicus* Cdc6/Orc1 homologues. The *M. thermautotrophicus* origin contains an AT-rich region and a number of repeated 13 bp sequence elements. *Sulfolobus sp.* encode three Cdc6/Orc1 homologues. By systematically analysing these three genes using two-dimensional electrophoresis, two active origins were identified in *Sulfolobus solfataricus* (Robinson *et al.*, 2004). One was adjacent to Cdc6-1 (oriC1) and the other was adjacent to Cdc6-3 (oriC2). Later, consistent results were obtained and a third origin (oriC3) was identified in a study using microarray-based marker frequency analysis (Lundgren *et al.*, 2004). OriC3 does not map next to a Cdc6/Orc1 gene. These three origins contain AT-rich regions flanked by various sequence repeats that were found to be binding sites for the Cdc6/Orc1 proteins. The repeated sequence elements identified in oriC1 were named origin recognition boxes (ORBs). Interestingly the second *S. solfataricus* origin (oriC2) contains repeats that are related to a core sequence found in those ORBs and were named mini-ORBs. ORB like elements have also been identified in the origins of *P. abyssii* and *Halobacterium* NRC-1 and mini-ORBs are equivalent to the 13 bp repeats found in the *M. thermautotrophicus* origin. In fact, *P. abyssii* and *Halobacterium* ORB elements can be recognised and bound by purified *S. solfataricus* Cdc6-1. These observations suggest that a similar mechanism of origin recognition is shared by a diverse range of archaeal organisms. Moreover, these conserved Cdc6/Orc1 binding sites are reminiscent of DnaA boxes in bacteria.

General features of archaeal origins appear to be an AT-rich stretch flanked by various inverted repeat elements in an intergenic region adjacent to a Cdc6/Orc1 gene. For instance, *S. solfataricus* oriC2 mini-ORB elements are inverted relative to one another and flank an AT-rich region containing the leading/lagging strand transition point. The orientation of the mini-ORBs may serve to impart directionality on the binding of the Cdc6/Orc1 proteins. Under supercoiling conditions such as those *in vivo*, it has been suggested that inverted repeats may lead to formation of

secondary DNA structure in origin regions (Kelman and Kelman, 2003). Interestingly, although the global architecture of archaeal origins resembles that of bacterial *oriCs*, the presence of multiple origins on a single chromosome is more akin to the eukaryotic mode of replication.

Cdc6

In the Archaea the best candidate for the initiator of DNA replication is a protein related to both Cdc6 and Orc1 in the Eukarya. Archaeal Cdc6/Orc1 proteins are believed to be involved in both origin recognition and helicase loading. Phylogenetic analysis of archaeal Cdc6/Orc1 homologues suggests that they can be divided into two distinct clades, Cdc6-1 and Cdc6-2 (Fig. 1-8).

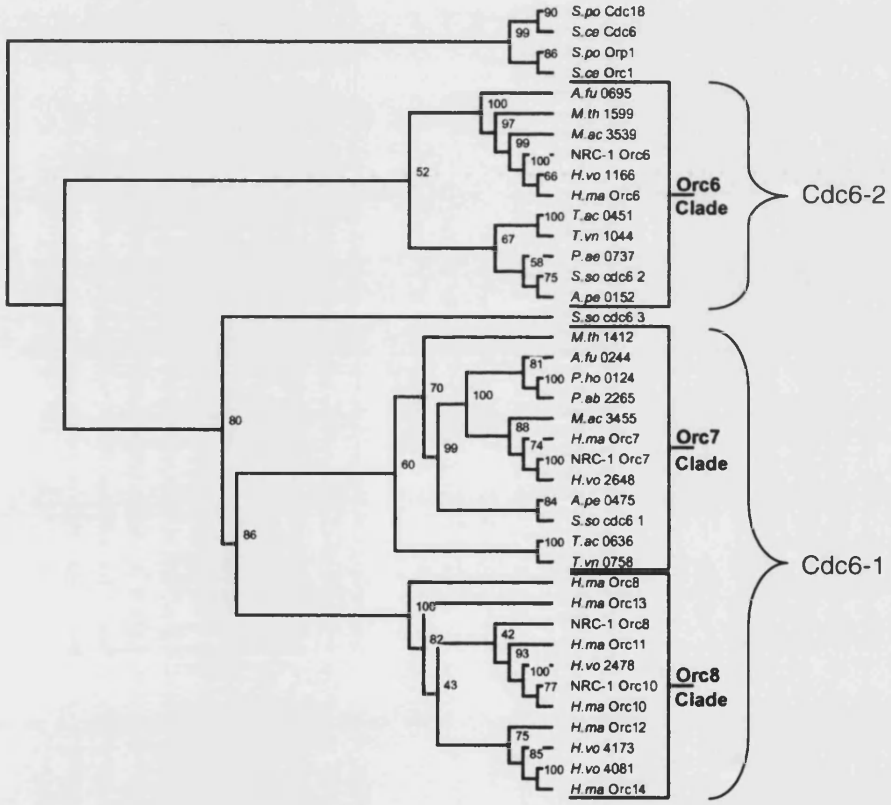


Figure 1-8 Neighbour-joining tree of Orc1-Cdc6 homologues found in sequenced archaeal genomes (Berquist and DasSarma, 2003)

It is not yet clear whether these two clades have distinct functions, or in fact structures, since both of the structures solved to date are in clade 2. Many archaeal species contain multiple Cdc6/Orc1 paralogues and, when that is the case, they encode at least one protein from each clade. Archaeal Cdc6/Orc1 proteins are

members of the ATPase Associated with various cellular Activities (AAA^+) superfamily. A phylogenetic study of the AAA^+ superfamily grouped the archaeal Cdc6/Orc1 homologues into a distinct clade that also contained DnaA, DnaC, and the eukaryal Cdc6 and ORC proteins (Iyer *et al.*, 2004). This may suggest a degree of functional conservation between the initiator proteins from each of the three domains. ATPase activity, at least, is a common feature. DnaA requires ATP binding for oligomerisation at *oriC* and ORC proteins also bind eukaryotic origins in an ATP dependent manner (Felczak and Kaguni, 2004, Bell and Stillman, 1992). Furthermore, the initiator proteins from all three domains share a modular architecture with an N-terminal AAA^+ domain linked to a C-terminal DNA-binding domain by a flexible tether. However, whereas the DNA-binding domain of DnaA is a helix-turn-helix (HTH), eukaryotic Orc1 and archaeal Cdc6/Orc1 have the related winged helix-turn-helix (WH) domain (Cunningham and Berger, 2005). Archaeal Cdc6/Orc1 proteins exhibit autophosphorylation activity on serine residues. Although the function of this is not yet clear, it appears to be modulated by interactions with DNA and MCM.

Cdc6-Origin Interaction

Several independent studies have demonstrated origin specific DNA binding by Cdc6/Orc1 proteins in archaea (Matsunaga *et al.*, 2001, Robinson *et al.*, 2004, Capaldi and Berger, 2004). Shortly after the *P. abyssii* origin was identified, chromatin immunoprecipitation studies indicated that the product of the single Cdc6/Orc1 gene was associated specifically with the origin *in vivo* (Matsunaga *et al.*, 2001).

M. thermautotrophicus Cdc6-1 and Cdc6-2 interact with DNA in different ways. *M. thermautotrophicus* contains two Cdc6/Orc1 homologues, one from the Cdc6-1 clade and one from the Cdc6-2 clade. Cdc6-1 specifically binds short 13 bp sequence elements (mini-ORBs), which are present throughout the origin, via its winged helix domain (Capaldi and Berger, 2004, Kasiviswanathan *et al.*, 2006). It is thought that it may bind cooperatively at mini-ORB elements and oligomerise in a manner similar to DnaA at *oriC*. In contrast, Cdc6-2 does not preferentially bind origin DNA over other sequences (Kasiviswanathan *et al.*, 2006). Furthermore, Cdc6-1 binds dsDNA but not ssDNA, whereas Cdc6-2 binds to both ss and dsDNA (Kasiviswanathan *et*

al., 2006). Replacing conserved residues in the helix-turn-helix portion of the winged helix domain of a Cdc6-1 protein dramatically reduces its affinity for origin DNA (Kasiviswanathan *et al.*, 2006, Robinson *et al.*, 2004, Capaldi and Berger, 2004). This suggests that proteins in the Cdc6-1 clade may recognise origin binding sites in a manner analogous to DnaA. It is tempting to speculate that they may also remodel the origin DNA in a manner analogous to DnaA. However, in Cdc6-2 proteins, conservation of residues in the wings rather than the helix-turn-helix suggests that they may bind DNA by a different manner. Furthermore, replacement of helix-turn-helix residues in *M. thermautotrophicus* Cdc6-2 did not completely abolish DNA binding as it did with equivalent residues in Cdc6-1 (Kasiviswanathan *et al.*, 2006). DNA inhibits the autophosphorylation activity of both Cdc6s, although Cdc6-1 to a lesser extent than Cdc6-2.

S. solfataricus Cdc6/Orc1 proteins may have different regulatory effects on initiation of DNA replication. *S. solfataricus* has three Cdc6/Orc1 homologues, one each from the Cdc6-1 and Cdc6-2 clades and a third (Cdc6-3) that is distantly related to Cdc6-1 (Fig. 1-8). Origin recognition appears to be mediated by the binding of distinct subsets of these proteins (Robinson *et al.*, 2004). They have distinct abundance profiles and their expression appears to be regulated temporally over the course of the cell cycle (Robinson *et al.*, 2004). Cdc6-1 and Cdc6-3 are present in the G1- and S-phases and bind cooperatively at the origins, suggesting that they promote DNA replication. Cdc6-2, however, is only expressed in G2 and so has been proposed to act as a negative regulator. This is supported by the observation that Cdc6-2 binds to sites that overlap with the Cdc6-1 and Cdc6-3 binding sites. This situation is reminiscent of SeqA a negative regulator of replication in *E. coli.* Biochemical studies did not detect any significant effect of ATP or ADP on the binding of Cdc6-1 to ORB elements (Robinson *et al.*, 2004). This is in contrast to eukaryotic ORC proteins. However, the relative position of the AAA^+ and winged helix domains in the *Aeropyrum pernix* Cdc6/Orc1 protein was influenced by the status of the nucleotide bound (Singleton *et al.*, 2004). Sequence analysis has failed to detect clear ORB elements in the genome of *Pyrobaculum aerophilum* (Robinson *et al.*, 2004). However, its single Cdc6/Orc1 homologue is in the Cdc6-2 clade, which may suggest a distinct mode of origin recognition.

MCM

MCM proteins in eukaryotes and archaea almost certainly function as the replicative DNA helicase complex. Eukaryotes have six different essential MCM proteins whereas many archaea have a single MCM homologue. The best characterised archaeal MCM protein is that of *M. thermautotrophicus*. *In vitro*, *M. thermautotrophicus* MCM exhibits DNA stimulated ATPase activity, binds to ssDNA and possesses DNA helicase activity that is ATP dependent and processive in a 3'-5' direction for up to 500 bp (Shechter *et al.*, 2000, Chong *et al.*, 2000, Kelman *et al.*, 1999). The protein exists as a large macromolecular complex with a molecular mass (~850 KDa) consistent with it forming a dodecamer (Chong *et al.*, 2000, Kelman *et al.*, 1999). This observation was supported when the crystal structure of the N-terminal portion of *M. thermautotrophicus* MCM revealed a barrel shaped double-hexamer architecture, with a long central channel capable of accommodating dsDNA (Fletcher *et al.*, 2003). The helicase activity resides in the complex rather than the monomer (Chong *et al.*, 2000). Furthermore, if a mutation is introduced that allows formation of single hexamers but not double hexamers then helicase activity is lost (Fletcher *et al.*, 2005). The MCM monomer consists of two main portions. The N-terminal portion is involved in protein multimerisation and DNA binding whereas the C-terminal portion contains the helicase catalytic domains. The crystal structure showed that the N-terminal portion folds into three distinct domains (A, B and C). Domain A at the N-terminus is mostly α-helical and domain B contains an important zinc-finger motif that participates in ssDNA binding (Kasiviswanathan *et al.*, 2004). Domain C connects the N-terminal portion of the enzyme to the C-terminal helicase portion and contains a β-finger involved in ss and dsDNA binding (Kasiviswanathan *et al.*, 2004). Domain C is also necessary and sufficient for MCM protein multimerisation and is required for interaction with Cdc6 (Kasiviswanathan *et al.*, 2004, Kasiviswanathan *et al.*, 2005). The quaternary arrangement of the *M. thermautotrophicus* MCM double-hexamer bares a striking resemblance to that of SV40 T-antigen. *M. thermautotrophicus* MCM also contains a β-hairpin in its AAA^+ domain similar to that found in T-antigen. Evidence suggests that the T-antigen uses a mode of unwinding DNA whereby encircled dsDNA is drawn into the enzyme by both opposed hexamers and the single DNA strands are extruded through gaps in the hexamer-hexamer interface (Fig. 1-9) (Duggin and Bell, 2006).

Figure 1-9 Model for SV40 T-antigen DNA interaction (Stillman, 2005)

Thus, the MCM double hexamer may function by a similar mode of action. In the C-terminal portion of the protein a β-α-β insert from helix-2 of the ATPase domain has been recently demonstrated to be important in transducing energy from ATP hydrolysis into duplex unwinding (Jenkinson and Chong, 2006). However, the helix-2 insert is unique to MCM suggesting a notable difference from the architecture of the T-antigen model. Studies show that *M. thermautotrophicus* MCM can also translocate along dsDNA and a suggested function of this is to displace proteins such as histones from DNA ahead of the replication fork (Stillman, 2005).

Cdc6-MCM Interaction

Using yeast two-hybrid analysis, it has been demonstrated that *M. thermautotrophicus* Cdc6-1 and MCM interact (Shin *et al.*, 2003, Kasiviswanathan *et al.*, 2005). However, no such interaction was detected, using this method, between Cdc6-2 and MCM or between Cdc6-1 and Cdc6-2 (Kasiviswanathan *et al.*, 2005). The interaction between Cdc6-1 and MCM was species specific, i.e. if either protein was replaced with a homologue from another species, no interaction was observed. Further investigation showed that the interaction is mediated predominantly via the winged helix domain of Cdc6-1 and domain C of the N-terminal portion of MCM (Kasiviswanathan *et al.*, 2005). In addition, DNA binding by Cdc6-1 was not needed for the Cdc6-1 MCM interaction. Further analysis involving far western and pull-down assays confirmed the two-hybrid results, but they also showed that Cdc6-2 and MCM do in fact interact (Kasiviswanathan *et al.*, 2005). However, this interaction is not via the winged helix domain as it is with Cdc6-1, rather, only full-length Cdc6-2 showed appreciable MCM binding. It was suggested that the interaction might be similar to that of eukaryotic Cdc6 and MCM, which is via the two proteins' AAA^+ domains.

Interaction with Cdc6 proteins inhibits the helicase and DNA translocation activities of MCM (Shin *et al.*, 2003, Kasiviswanathan *et al.*, 2005). This is an interesting

observation because it is reminiscent of the inhibition of DnaB helicase by DnaC, the helicase loader in bacteria. This effect is dependent on ATP-binding by Cdc6-1, but not on ATP hydrolysis or DNA binding. While Cdc6-1 and Cdc6-2 both inhibit MCM helicase activity, Cdc6-2 does so much more effectively.

The autophosphorylation activities of Cdc6-1 and Cdc6-2 are modulated by MCM (Kasiviswanathan *et al.*, 2005). However, whereas Cdc6-1 autophosphorylation is stimulated by MCM, Cdc6-2 autophosphorylation is strongly inhibited. The presence of MCM substantially reduces DNA binding by Cdc6-1 and Cdc6-2 *in vitro* (Kasiviswanathan *et al.*, 2006). Thus, it has been suggested that MCM and DNA may compete for Cdc6 binding.

It has been hypothesised that Cdc6-1 and Cdc6-2 have different roles in initiation of DNA replication. It was proposed that Cdc6-1 may be a functional homologue of eukaryotic ORC or bacterial DnaA, involved in origin recognition. The Cdc6-1 gene is adjacent to the *M. thermautotrophicus* origin and it exhibits preferential binding to origin repeats over random sequences (Capaldi and Berger, 2004). Cdc6-2, on the other hand, was proposed to be a functional homologue of eukaryotic Cdc6 or bacterial DnaC, involved in loading the helicase. Cdc6-2 inhibits MCM helicase activity more that Cdc6-1 (just as DnaC inhibits DnaB) and it has no preference for origin DNA over random sequences (Kasiviswanathan *et al.*, 2005). This hypothesis is supported by the differential modulation of autophosphorylation of the two proteins by MCM. Is the hypothesis also true for members of the two Cdc6 clades in species other that *M. thermautotrophicus*?

Other Insights into DNA Replication in Archaea

Archaea utilize B-type and archaea specific D-type DNA polymerases. All sequenced archaea encode between one and three B-type DNA polymerases. They are similar to PolB in eukarya and bacteria and they possess a 3'-5' exonuclease proofreading activity. They are good candidates for the replicative enzymes and are the only polymerases found in some crenarchaeal species. Euryarchaeota, however, also all contain PolD, a euryarchaeota specific DNA polymerase. PolD is a dimer consisting of a large catalytic subunit (D2) and a small accessory subunit (D1). D-type polymerases exhibit proof reading activity. In euryarchaea PolB and PolD may

both participate in DNA replication; one may replicate the leading strand while the other replicates the lagging strand. However, PolD interacts with the Rad51 homologue perhaps suggesting a role in DNA recombination (Hayashi *et al.*, 1999). Archaea contain archaea-specific DNA polymerases and archaea-specific adaptations to more common enzymes. Further studies are required to investigate the roles of these features in archaeal DNA replication.

The archaeal sliding clamp is homologous to eukaryotic PCNA. PCNA complexes are ring shaped trimers that encircle dsDNA behind the polymerase, tethering it to the DNA and increasing its processivity. In euryarchaeal species, as in most eukaryotes, a single PCNA homologue has been identified. For example, the PCNA homologue in *M. thermautotrophicus* is believed to form a homotrimer. In each crenarchaeal species, however, three homologues have been identified and are thought to have arisen from early lineage-specific gene duplication. The reasons for there being multiple homologues are not yet known. However, the three PCNA homologues in *S. solfataricus* form a heterotrimer and each subunit interacts with different other components of the replication machinery (Dionne *et al.*, 2003). For example PCNA2 interacts with PolB1 whereas PCNA1 and 3 interact with FEN1 and DNA ligase 1, enzymes involved in replacing RNA primers with DNA (Dionne *et al.*, 2003). Interestingly, PCNA1 also interacts with a uracil glycosylase, an enzyme that repairs lesions caused by base deamination (Dionne and Bell, 2005). The archaeal sliding clamp is homologous to that of Eukarya, though having a heteromeric PCNA is a trait specific to archaea. Interestingly, while heteromeric eukaryotic complexes tend to be replaced by homomeric complexes in archaea, with PCNA the trend is reversed.

1.4 AIMS

Research into the initiation of DNA replication in archaea has revealed some interesting molecular details. For instance, some species appear to initiate replication from a single origin (e.g., *M. thermautotrophicus*), whereas others have multiple origins (e.g., *S. solfataricus*). Moreover, clear homologues of the eukaryotic initiation proteins MCM and Cdc6 have been identified in almost every archaeal species for which the genome sequence is known. However, the molecular mechanism(s) archaea use to regulate the initiation of DNA replication with respect to the cell cycle remains almost completely unknown. Therefore, the aim of this project was to address this deficit by identifying archaeal proteins that interact with an archaeal *oriC*, MCM and Cdc6. This approach was based on several premises: first, a system for controlling initiation exists in archaea. Second, the *oriC* element and the MCM and Cdc6 initiation proteins are involved in this system. And lastly, other, as yet unidentified, components of this system interact directly with *oriC*, MCM or Cdc6. Therefore, the yeast one-hybrid system was used to screen a *M. thermautotrophicus* genomic shot-gun library for proteins that interact with the *oriC* from that species. In addition, the yeast two-hybrid system was used to screen the same library for proteins that interacted with MCM and Cdc6-1 from *M. thermautotrophicus*. In the case of one putative interacting protein (interaction with Cdc6-1), preliminary biochemical characterisation of NTPase activity was carried out.

2 Materials and Methods

2.1 Media and Solutions

2.1.1 Media

- Agar Plates: appropriate media and 1.6% (w/v) agar (bacteriological grade)
- Luria-Bertani (LB): 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl
- Synthetic Drop-out (SD) media: 0.67% (w/v) Difco yeast nitrogen base (without amino acids), 2% (w/v) dextrose and the following:
    - L-adenine hemisulfate 20 mg/L
    - L-arginine HCl 20 mg/L
    - L-histidine HCl monohydrate 20 mg/L
    - L-isoleucine 30 mg/L
    - L-leucine 100 mg/L
    - L-lysine HCl 30 mg/L
    - L-methionine 20 mg/L
    - L-phenylalanine 50 mg/L
    - L-threonine 200 mg/L
    - L-tryptophan 20 mg/L
    - L-tyrosine 30 mg/L
    - L-uracil 20 mg/L
    - L-valine 150 mg/L
    - One or more of these nutrients was omitted to obtain the desired drop-out medium. For example, L-tryptophan and L-leucine were omitted to obtain SD/–Trp/–Leu medium.
- SOB: 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.05% (w/v) NaCl, 2.5 mM KCl and 10 mM $MgCl_2$.
- SOC: as SOB but supplemented with 20 mM glucose.
- YPAD: 2% (w/v) peptone, 1% (w/v) yeast extract, 2% (w/v) dextrose, 30 mg/L adenine hemisulfate.

2.1.2 Solutions

- Coomassie blue: 0.25% (w/v) Brilliant Blue R 250, 45% (v/v) methanol and 10% (v/v) acetic acid
- Coomassie destain: 20% (v/v) methanol and 10% (v/v) acetic acid
- GF buffer: 20 mM Tris-HCl, pH 7.9, 1 M NaCl and passed through a filter with 0.22 μm pore size
- Inoue transformation buffer: 55 mM $MnCl_2$, 15 mM $CaCl_2$, 250 mM KCl, 10 mM PIPES, adjusted to pH 6.7 with KOH and filter sterilised
- Lyticase enzyme stock: 20 Units/μl lyticase (Sigma), 10 mM $NaPO_4$, pH 7.4 and 50% (v/v) glycerol
- Malachite green reagent: 0.03% (w/v) malachite green oxaolate, 8.3 mM Na molybdate, 0.7 M HCl and 0.05% (v/v) Triton X-100
- Phenol:chloroform: 1 volume of TE washed phenol to 1 volume chloroform
- Proteinase K: 0.2 mg/ml PCR grade proteinase K (Roche), 10 mM Tris-HCl, pH 8.5 and 0.1 mM EDTA
- TAE: 40 mM Tris-acetate, pH 8.5 and 1 mM EDTA
- TBS-T: 10 mM Tris-HCl, pH 7.4, 0.9% (w/v) NaCl and 0.1% (v/v) Tween-20
- TE: 10 mM Tris-HCl, pH 8.5 and 1 mM EDTA
- T0.1E: 10 mM Tris-HCl, pH 8.5 and 0.1 mM EDTA
- Yeast lysis buffer: 50 mM Tris-HCl, pH 7.5, 1.2 M sorbitol, 10 mM EDTA, and 100 mM β-mercaptoethanol

2.2 Strains, Vectors and Oligonucleotides

2.2.1 Strains

E. coli

- DH5α
  - o F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ–
- XL10 GOLD
  - o endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet^R F'[proAB $lacI^qZ$ΔM15 Tn10(Tet^R Amy Cm^R)]

- BL21 (DE3)
    - F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3)
- BL21 (DE3) pLysS
    - F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3) pLysS(cm^R)
- BL21 (DE3) Star Rosetta 2
    - F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ rne131 λ(DE3) pRARE2(cm^R)

S. cerevisiae

- AH109
    - *MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2 : : $GAL1_{UAS}$-$GAL1_{TATA}$-HIS3, $GAL2_{UAS}$-$GAL2_{TATA}$-ADE2, URA3 : : $MEL1_{UAS}$-$MEL1_{TATA}$-lacZ*
- YM4271
    - *MATa, ura3-52, his3-200, ade2-101lys2-801, leu2-3, 112, trp1-901, tyr1-501, gal4-Δ512, gal80-Δ538, ade5 : : hisG*

2.2.2 Vectors

- pGEM-T Easy (Promega)
    - Designed for cloning PCR amplicons by utilising single 3′-T overhangs (for vector map see Section 10.1.1).
- pGBKT7 (Clontech)
    - pGBKT7 is the DNA-binding domain (DNA-BD) bait vector included with MATCHMAKER Two-Hybrid System 3. The multiple cloning site (MCS) of pGBKT7 contains unique restriction sites for constructing fusion proteins with GAL4 DNA-BD (Section 10.1.2).
- pGADT7 (Clontech)
    - pGADT7 is the activation domain (AD) prey vector included with MATCHMAKER Two-Hybrid System 3. The MCS of pGADT7 contains unique restriction sites for constructing fusion proteins with the GAL4 AD and either a protein of interest or a fusion protein library (Section 10.1.3).
- pHisi-1 (Clontech)

- pHISi-1 is a yeast integration and reporter vector for use with the MATCHMAKER One-Hybrid System. pHISi-1 contains the yeast *HIS3* gene downstream of the MCS (Section 10.1.4).

- pET28a (Novagen)
  - pET28a is an *E. coli* T7-lac protein expression vector.
  - pET28a carries an N-terminal His-tag/thrombin/T7-Tag configuration; plus an optional C-terminal His-tag sequence (Section 10.1.5).
- pCDFDuet-1 (Novagen)
  - Designed for the co-expression of two target ORFs in *E. coli*, one with an N-terminal His-tag and one with a C-terminal S-tag (10.1.6).

2.2.3 Oligonucleotides

Name	**Sequence (5'-3')**
OB13	tcaggttcacgcttaacc
OB14	aaactgctgcatcaaatttcag
Mth1888EcoR1for	ggaattcggccgattaggttgttggtaggg
Mth1888EcoR1rev	ttccgaattcaagagggcgggttacagccc
Mth1412start	ggtcatatgaacatttttgatgagataggg
Mth1412stop	aaaggatccttaaacaccccagagtgagtc
Mth1770start	cttcatatgatgaaaaccgtggataagagc
Mth1770stop	atgggatcctcagactatcttaaggtatcc
Mth810startNde1	gaggcgtgttcatatgaagtccctcccacc
Mth810stopPst1	gccattatctgcagatcagaacctttcacagtacc
C1412startBamH1	tgaagggatcctatgaacatttttgatgagataggg
C1412stopHind3	gaaaaaagctttaaacaccccagagtgagtcctcc
C203startNde1	gaagatcatatgaaaggattagaatttagtgag
C203endXho1	aaaaatctcgagtttggatgattccagtacatcc
C810endXho1	gaggctcgaggaacctttcacagtaccttc
C1770endXho1	taaatgctcgaggactatcttaaggtatcccc
Mth203startNhe1	ggaagctagcatgaaaggattagaatttagtgag
Mth203stopHind3	ctcaaaagcttattatttggatgattccagtacatcc
pGADT7for	actcactatagggcgagcgcc
pGADT7rev	cacagttgaagtgaacttgcgggg

2.3 Standard Methods

2.3.1 Transformation of *E. coli*

Competent *E. coli* strains were prepared and transformed according to the method by Inoue and colleagues as described in Sambrook and Russell (Inoue *et al.*, 1990, Sambrook and Russell, 2001). Typically, a single fresh *E. coli* colony was used to inoculate 25 ml of SOB medium in a 250 ml conical flask. This starter culture was incubated for 6 to 8 hrs at 37°C with shaking (200 rpm). A 1 L conical flask containing 250 ml of SOB was then inoculated with 10 ml of the starter culture and incubated overnight at room temperature (typically 18-22°C), with shaking (200 rpm) until the OD_{600} reached 0.55. The flask was then cooled on ice for 10 min and the cells were harvested by centrifugation at 2,500 *g* for 10 min at 4°C. Next, the cells were washed with 80 ml of ice-cold Inoue transformation buffer (55 mM $MnCl_2$, 15 mM $CaCl_2$, 250 mM KCl, 10 mM PIPES, adjusted to pH 6.7 with KOH and filter sterilised) and centrifuged as above. The cells were then resuspended in 20 ml of ice-cold Inoue buffer, supplemented with 1.5 ml dimethylsulfoxide (DMSO) and incubated on ice for 10 min. 100 µl aliquots of this cell suspension were then snap-frozen in liquid nitrogen and stored at -80°C until needed.

For each transformation reaction, including positive and negative transformation controls, a 100 µl aliquot of competent *E. coli* was thawed on ice for 10 min and transferred to a 15 ml polypropylene tube. The transforming DNA was added and incubated with the cells, on ice, for a further 30 min. The tube was immersed in a 42°C water bath for 90 sec then briefly cooled on ice. 800 µl of SOC medium was added and the cells incubated for 45 min at 37°C with shaking (200 rpm) to allow expression of the plasmid encoded antibiotic resistance marker. Up to 200 µl of cell suspension was then spread onto a SOB agar plate (containing appropriate antibiotic) and incubated overnight at 37°C. Transformation efficiencies were typically in the order of 1×10^8 colony forming units (CFU)/µg of plasmid DNA.

2.3.2 Transformation of *S. cerevisiae*

Strains of *S. cerevisiae* were transformed according to the method developed by Gietz and Woods (Gietz and Woods, 2002). A starter culture was prepared as follows: a fresh yeast colony was used to inoculate 15 ml of YPAD (or drop-out

medium if it was necessary to maintain selection for a plasmid) and incubated overnight at 30°C with shaking (200 rpm). The starter culture was titered by measuring the OD_{600}. The cell density in cells/ml was estimated by referring to data from a previous yeast OD_{600} calibration experiment. A volume containing 2.5×10^8 cells was then added to 60 ml of YPAD (or drop-out medium) pre-warmed to 30°C. The culture was incubated at 30°C with shaking (200 rpm) until an OD_{600} of 2.5 was reached (measured by diluting the culture), which corresponded to a density of 2×10^7 cells/ml. 50 ml of culture was harvested by centrifugation at 1,000 *g* and room temperature for 5 min. The cells were washed once with 25 ml of sterile milliQ then resuspended in sterile milliQ to give a final volume of 1 ml. 100 µl aliquots (each containing *ca.* 10^8 cells) were transferred to 1.5 ml microfuge tubes, centrifuged at 16,000 *g* for 30 sec and the supernatant removed. For each transformation, one pellet of cells was resuspended in 360 µl of transformation mix containing: 1 µg of the plasmid to be transformed, 100 mM lithium acetate, 33% (w/v) PEG 3350 and 0.28 mg/ml salmon sperm carrier DNA (boiled for 5 min then cooled on ice just before use). The tubes were incubated in a 42°C water bath for 50 min then centrifuged at 16,000 *g* for 30 sec. The supernatant was removed and the cells resuspended in 1 ml of sterile milliQ. Cell suspension was spread onto plates containing the appropriate drop-out medium (e.g., SD/-Trp/-Leu), and incubated for 4 days at 30°C. Positive and negative transformation controls were carried out in parallel with experimental samples. Transformation efficiencies were typically in the order of 10^4 CFU/µg of plasmid/10^8 cells.

2.3.3 Purification of plasmid DNA (*E. coli* and *S. cerevisiae*)

E. coli

A single colony of *E. coli* DH5α harbouring the relevant plasmid was used to inoculate 5 ml of LB containing an appropriate antibiotic in a sterile 50 ml tube. The culture was incubated overnight at 37°C with shaking (200 rpm). 2 ml of the culture was centrifuged at 16,000 *g* for 2 min at room temperature and the supernatant removed. The pellet of cells was then processed using a QIAprep Spin Miniprep Kit (Qiagen), according to the manufacturer's instructions. Plasmids were typically eluted in 50 µl of TE or 10 mM Tris-HCl, pH 8.5.

S. cerevisiae

The strain of yeast harbouring the relevant plasmid was used to inoculate 5 ml of the appropriate drop-out medium (e.g., SD/-Trp/-Leu) in a sterile 50 ml tube. The culture was incubated overnight at 30°C with shaking (200 rpm). 2 ml of culture was centrifuged at 16,000 *g* for 1 min at room temperature and the supernatant removed. The cells were resuspended in 50 µl of yeast lysis buffer (50 mM Tris-HCl pH 7.5, 1.2 M sorbitol, 10 mM EDTA, and 100 mM β-mercaptoethanol). The yeast cell walls were digested by adding 10 µl of lyticase enzyme stock (20 Units/µl lyticase [Sigma], 10 mM $NaPO_4$ [pH 7.4], 50% [v/v] glycerol) followed by incubation overnight at 37°C. The resultant spheroplasts were collected by centrifugation at 1,500 *g* for 5 min at room temperature. The supernatant was removed and the pellet was processed using a QIAprep Spin Miniprep Kit, according to the manufacturer's instructions. Because of the very low yield of plasmids from yeast, the resulting plasmid DNA was amplified by firstly transforming into DH5α (2.3.1), and then purified as above (2.3.3).

2.3.4 Purification of PCR amplicons and DNA from agarose gels

Products from preparative PCRs (typically a 100 µl reaction) were purified for cloning using a QIAquick PCR purification kit (Qiagen), according to the manufacturer's instructions. DNA fragments, to be used for cloning, were isolated from agarose gel slices with a QIAquick gel extraction kit (Qiagen), according to the manufacturer's instructions.

2.3.5 Ethanol precipitation of DNA

DNA was precipitated by adding sodium acetate, pH 5.2, to 0.3 M [final] then 3 volumes of 100% (v/v) ethanol, followed by incubation at -20°C for 30 min or longer. The precipitated DNA was collected by centrifugation at 16,000 *g* for 30 min at room temperature. The pellet was washed with 70% (v/v) ethanol and dried under vacuum before being resuspended in a suitable buffer.

2.3.6 DNA electrophoresis

DNA resolving gels were made by dissolving agarose (Fisher, typically 0.8, 1.0 or 1.5% [w/v]) in 1 × TAE buffer (40 mM Tris-acetate, pH 8.5, 1 mM EDTA) and casting it in a standard gel tray. DNA samples were prepared for electrophoresis by

addition of Blue Orange 6 × loading dye (Promega) to a final concentration of 1 ×. The samples were run along side molecular mass markers (typically 1 Kb DNA Ladder, Promega) at a constant voltage (typically 100 V) for a time necessary to give adequate seperation of the bands. After electrophoresis, gels were stained in a solution of 1 μg/ml Ethidium Bromide in 1 × TAE, destained in water and photographed on a UV transilluminator.

2.3.7 Measurement of DNA concentration

The concentration of a solution of DNA was measured by diluting a small sample (typically 10 μl) in milliQ water (typically 790 μl) and reading the A_{260} using a quartz cuvette (1 cm path length). Then the following formula was used to calculate the concentration in μg/μl:

$[A]\ \mu g/\mu l = A_{260} \times Z \times ([X + Y] \times 10^{-3}) / X$

Where: A = concentration of original solution, Z = 50 for dsDNA or 33 for ssDNA, X = volume of DNA sample (μl) and Y = volume of water (μl).

2.3.8 PCR

M. thermautotrophicus genes were amplified by PCR using primers designed to facilitate their cloning into the target vector. The following considerations were taken into account when designing primers. Restriction sites were designed into the 5' end of the forward and reverse primers. When digested with the appropriate restriction enzymes and ligated into the target expression vector, the restriction sites were designed to place the gene in frame with the target vector's protein coding sequence. The target gene's DNA sequence was checked to make sure that these restriction sites were not present within it. When possible, different restriction sites were used in the forward and reverse primers to enable directional cloning. Each primer was designed to have as similar a melting temperature (T_M) to its partner as possible. Primers were typically 30-36 bases in length and had 3' GC clamps when possible.

For each preparative PCR, a 100 μl reaction was set up on ice. The reaction composition was as follows: 1.5 ng/μl *M. thermautotrophicus* genomic DNA (prepared as described in Section 2.6.1), 400 nM of each of the forward and reverse primers (MWG), 250 μM of each dNTP (Promega), either 0.03 Units/μl of Pfu

(Promega) or 0.02 Units/µl of Phusion (NEB) DNA polymerase and either 1 × Pfu or Phusion reaction buffer.

M. thermautotrophicus genes were amplified by touchdown PCR in an Eppendorf Mastercycler. The conditions used to amplify a specific gene were based on the following generic PCR program:

95°C	5 min	
95°C	1 min	× 10 cycles with T_A decreasing 1°C per cycle from (T_A FINAL + 9°C) to (T_A FINAL).
T_A°C	1 min	
72°C	E min	
95°C	1 min	× 28 cycles
T_A FINAL°C	1 min	
72°C	E min	
72°C	10 min	

The annealing temperature 'T_A FINAL' was typically 7°C lower than the lowest primer T_M of the pair, and the elongation time 'E' was the length of the gene to be amplified in kilobases × 2 min (for Pfu) or × 1 min (for Phusion). These PCR programs were named according to the values of T_A FINAL°C and time E min used, e.g., 57°C / 1.5 min.

2.3.9 Colony PCR (*E. coli* and *S. cerevisiae*)

E. coli

After performing a cloning step, PCR was used to identify colonies that contained the desired construct. The primers employed for this were either those used in the initial PCR or ones directed against the vector, flanking the MCS. Each colony was picked with a pipette tip and spotted first onto a LB agar plate (with appropriated antibiotic), and the residual cells were then deposited into a PCR tube containing 25 µl of proteinase K solution (10 mM Tris-HCl, pH 8.5, 0.1 mM EDTA and 0.2 mg/ml PCR grade proteinase K [Roche]). This was incubated for 15 min at 55°C to lyse the cells, and then 15 min at 80°C to inactivate the proteinase K. 1 µl of this lysate was used as template in a 15 µl PCR containing additionally: 400 nM each primer, 250

μM dNTPs, 0.05 Units/μl Taq DNA polymerase (Promega) and 1 × GoTaq buffer (Promega). The test reactions were always set up alongside negative (no template) and positive (known template) controls. The reactions were incubated according to appropriate cycling conditions for primers and template. Agarose gels were run to check the success of the PCRs and the size of the products. Colonies that contained the desired construct were used to prepare glycerol stocks and plasmid minipreps.

S. cerevisiae

S. cerevisiae colony PCR was used to amplify library insert DNA from pGADT7 after the yeast one- or two-hybrid screens. pGADT7 forward and reverse primers were used (2.2.3). Each colony was picked with a pipette tip then spotted onto a SD agar plate (the plate was later incubated at 30°C.) The residual cells on the tip were then deposited into an Eppendorf containing 25 μl of yeast lysis buffer. The yeast cell walls were digested by adding 2 μl of lyticase enzyme stock followed by a 2 hr incubation at 37°C. After that, 900 μl of sterile milliQ was added and the tubes were boiled for 5 min to lyse the cells and release the plasmid. 0.5 μl of this lysate was used in a 15 μl PCR as described in the previous section. As before, agarose gels were run to check the size of the products.

2.3.10 PCR amplicon cloning

M. thermautotrophicus genes were amplified by PCR, cloned first into pGEM-T Easy vector (Promega) and then sub-cloned into the target vector. This two-step strategy was employed to facilitate restriction digestion and cloning into the target vector. A PCR amplicon was first A-tailed by combining the following in a 10 μl reaction: approximately 8.5 μM PCR amplicon, 200 μM dATP, 5 mM $MgCl_2$, 0.5 Units/μl Taq (Promega) and 1 × Taq reaction buffer. This was incubated for 20 min at 70°C. The A-tailed PCR amplicon was ligated into pGEM-T Easy by combining the following in 10 μl: 2.5 μM pGEM-T Easy, approximately 2.5 μM A-tailed PCR amplicon (directly from A-tail reaction), 0.3 Units/μl T4 DNA ligase (Promega) and 1 × Rapid ligation buffer (Promega). pGEM-T Easy ligation reactions were incubated at 4°C overnight. Positive and negative ligation controls were performed in parallel. The ligations were transformed into competent DH5α cells (Section 2.4.1). The transformed cells were plated onto SOB agar supplemented with 100 μg/ml ampicillin, 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 100

μg/ml 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal) for blue-white colony selection. Colony PCR was used to check white colonies for the presence of the correct insert.

2.3.11 Sub-cloning

Each gene was cut out of pGEM-T Easy using the restriction sites designed into the original primers. Restriction enzymes were from Promega. When different restriction sites were used, double digests were performed according to the manufacturer's directions. A typical double digest was prepared in 50 μl as follows: 0.1 μg/μl vector-clone DNA, 0.1 μg/μl bovine serum albumin (BSA), 0.2 Units/μl restriction enzyme A, 0.2 Units/μl restriction enzyme B and 1 × suitable buffer. The digestion was incubated for 2 hrs at 37°C then 20 min at 65°C (to inactivate the enzymes). Insert DNA was separated from vector by agarose gel electrophoresis and cleaned up with a gel extraction kit (2.3.4). The recipient vector was digested in the same way except that, shrimp alkaline phosphatase (SAP, Promega) was added after the 2 hr 37°C incubation to give 0.05 Units / pmol of cut DNA ends [final]; the reaction was then incubated for a further 30 min at 37°C before proceeding to the 65°C inactivation step. This SAP step was included to improve cloning efficiency by preventing relegation of any residual singly cut vector. Cut vector was then isolated from excised linker DNA by agarose gel electrophoresis and cleaned up (2.3.4). Vector and insert were ligated in a 10 μl reaction containing: 5 ng/μl of vector, an equimolar (approximately) concentration of insert, 0.3 Units/μl of T4 DNA ligase and 1 × Rapid ligation buffer. Negative control (no insert) reactions were set up in parallel and ligations were incubated overnight at 4°C. Ligation products were transformed into DH5α cells (2.3.1) and colony PCR (2.3.9) was used to confirm the presence of the correct insert. Positive clones were chosen to be sequenced to check for errors.

2.3.12 DNA sequencing

Plasmid DNA was prepared by Miniprep kit (2.3.3) and analysed by A_{260} and electrophoresis (2.3.7 and 2.3.6). For each sample, 1.5 μg of DNA was dried down in vacuum and sent to MWG for sequencing.

2.3.13 Protein electrophoresis

Protein samples were analysed by SDS-PAGE according to the method of Laemmli (1970) using the Mini-Protean 3 electrophoresis system (Bio-Rad). The resolving gel matrix was prepared with the following composition: 0.375 M Tris-HCl, pH 8.9, 0.1% (w/v) SDS, acrylamide (10% [w/v] T and 0.87% [w/v] C), 0.0417% (w/v) ammonium persulfate (APS) and 0.105% (v/v) N, N, N', N'-tetramethylethylenediamine (TEMED). The APS and TEMED were added just before the gel was cast; then it was over-laid with water and left to set for 30 min at room temperature. The stacking gel was prepared as follows: 0.166 M Tris-HCl, pH 6.8, 0.14% (w/v) SDS, acrylamide (4% T [w/v] and 0.35% [w/v] C), 0.0725% (w/v) APS and 0.145% (v/v) TEMED. Again, the APS and TEMED were added just before pouring the stacking gel, inserting the comb and leaving it to set for 30 min. Once set, gels were loaded into the tank and both electrode reservoirs filled with running buffer: 52 mM Tris, pH 8.3, 0.1% (w/v) SDS and 0.4% (w/v) glycine. A protein sample was prepared for SDS-PAGE by adding an equal volume of 2 × sample buffer (100 mM Tris-HCl, pH 6.8, 4% [w/v] SDS, 0.02% [w/v] bromophenol blue, 20% [v/v] glycerol and 0.7 M β-mercaptoethanol) and boiling it for 3 min. Up to 20 µl of sample was loaded per well, including a lane for protein molecular mass markers. Gels were run at 200 V for 50-70 min. Once a gel had finished running it was rinsed in distilled water and stained for 15 min in Coomassie blue solution (0.25% [w/v] Brilliant Blue R 250, 45% [v/v] methanol and 10% [v/v] acetic acid). The gel was then left overnight in Coomassie destaining solution (20% [v/v] methanol and 10% [v/v] acetic acid).

2.3.14 Measurement of protein concentration

A sample's protein concentration was measured using Bradford assay reagent (Bio-Rad) according to the manufacturer's instructions. A fresh standard curve, using bovine serum albumen (BSA), was prepared each time the assay was performed.

2.4 Preparation of Bait Constructs

The *M. thermautorphicus oriC* region and *mth1888* gene were cloned into pHisi-1 to form bait constructs for use in the yeast one-hybrid system. *oriC* was amplified by PCR (60°C / 1 min Phusion) with OB13 and OB14 primers and blunt-cloned directly into the SmaI site of pHisi-1. *mth1888* was amplified (62°C / 1.25 min Phusion)

with Mth1888EcoR1for and Mth1888EcoR1rev primers and cloned into the EcoR1 site of pHisi-1.

- pHisi-1/*oriC*
- pHisi-1/*mth1888*

cdc6-1, *mcm* and *mth810* were cloned into pGBKT7 to encode baits for the yeast two-hybrid system. *cdc6-1* was amplified by PCR (57°C / 2.5 min Pfu) with Mth1412start and Mth1412stop primers, cloned into pGBKT7 using Nde1 and BamH1 sites designed into the primers. *mcm* was amplified (57°C / 4.25 min Pfu) with Mth1770start and Mth1770stop primers and cloned into pGBKT7 using Nde1 and BamH1. *mth810* was amplified (60°C / 2.25 min Phusion) with Mth810startNde1 and Mth810Pst1 primers and cloned into pGBKT7 using Nde1 and Pst1.

- pGBKT7/*cdc6-1*
- pGBKT7/*mcm*
- pGBKT7/*mth810*

2.5 Preparation of *M. thermautotrophicus* Genomic DNA Library

2.5.1 Preparation of *M. thermautotrophicus* genomic DNA

M. thermautotrophicus genomic DNA was prepared according to a method originally developed by Dr. James Chong (University of York, UK). 15 g of *M. thermautotrophicus* cell mass was kindly provided by Dr. Alan Majerník (University of York, UK) and stored (-80°C) until needed. The cells were resuspended in 2 volumes of TE sucrose solution (20 mM Tris-HCl, pH 8.0, 5 mM EDTA and 10% [w/v] sucrose). The cell suspension was frozen with liquid nitrogen, and ground in a pestle and mortar. The viscous thawed lysate was then transferred to a 50 ml centrifuge tube, SDS (1.67% [w/v] [final]) was added, and it was incubated for 30 min at 60°C. Next, NaCl (0.685 M [final]) was added and the sample was incubated for a further 40 min on ice. At this point the sample was centrifuged at 20,000 *g* for 25 min at 4°C to pellet the cell debris. The supernatant was then decanted off and retained. It was then supplemented with isopropanol (33% [v/v] [final]) and incubated for 5 min at room temperature. The precipitated DNA was centrifuged at 18,000 *g* for 20 min at 4°C. The supernatant was discarded and the DNA pellet was

allowed to air dry before being resuspended in TE. RNase A (Roche) (0.1 mg/ml [final]) was added to the sample, which was then incubated for 30 min at 37°C. The DNA solution was extracted 3 times with 1:1 phenol:chloroform. Next, the DNA sample was used to dissolve 10 g of CsCl and then made up to 10 ml to give a 100% (w/v) CsCl solution [final]. Hoescht H33258 dye (Sigma) (0.15 mg/ml [final]) was added to the sample and it was centrifuged at 276,349 *g* (65,000 rpm Beckman Ti75 rotor) for 36 hours at 18°C. The DNA band was recovered (~1.5 ml) using a syringe with a 16G needle. To remove the Hoescht dye, the DNA was extracted five times with butanol (saturated with 100% (w/v) CsCl in water). The DNA was then dialysed overnight in 2 l of TE buffer. To concentrate it, the DNA was ethanol precipitated (2.3.5), resuspended in TE and stored at -20°C.

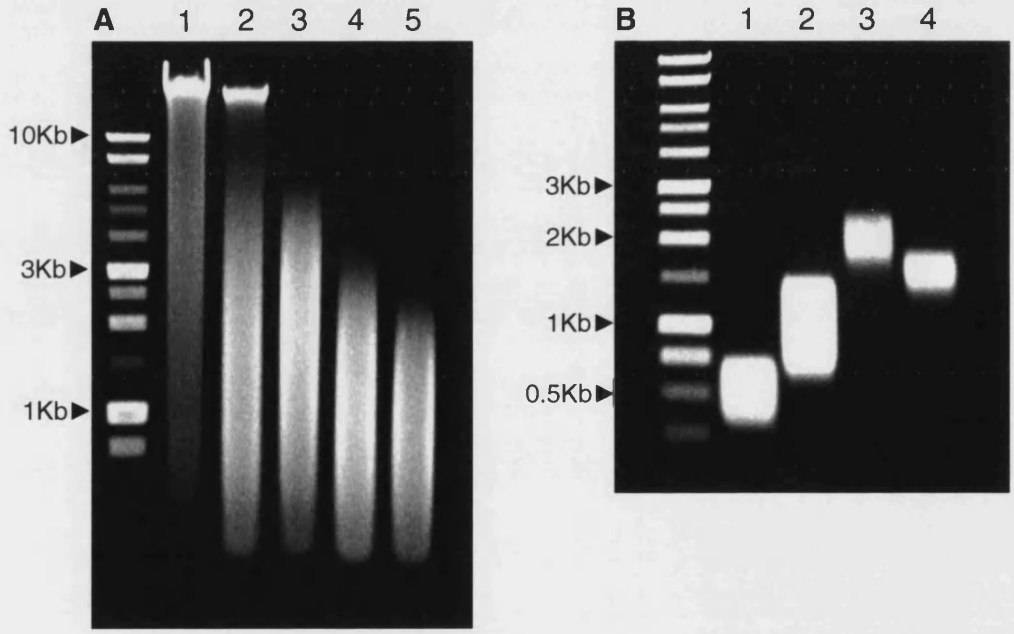


Figure 2-1 Preparation of *M. thermautotrophicus* genomic DNA for library cloning. A. Genomic DNA sheared by sonication (0.8% gel). 1) Unsonicated, 2) 10 sec, 3) 20 sec, 4) 30 sec and 5) 40 sec. B. Size fractionated sheared genomic DNA (0.8% gel). 1) 0.4-0.75 Kb, 2) 0.75-1.5 Kb, 3) 2-2.5 Kb and 4) 1.5-2 Kb.

2.5.2 Preparation of randomly sheared insert DNA

The *M. thermautotrophicus* DNA was sheared by using a Heat-Systems W380 Ultrasonic cup-horn sonicator. Sonication was chosen in order to obtain a completely random distribution of fragments. A trial was carried out to determine

the optimum conditions for sonication (Fig. 2-1A). Based on this trial, the genomic DNA was sonicated for 3 × 10 sec. 100 µg of sheared DNA was ethanol precipitated and resuspended in 120 µl of 1 × mung bean nuclease buffer (NEB). Mung bean nuclease (NEB) was added to give 1.25 Units/µl [final] and the reaction was incubated for 10 min at 30°C to digest away any single stranded overhangs. The DNA was then size fractionated by agarose gel electrophoresis as follows. The DNA was ethanol precipitated, resuspended in 1 × loading buffer and run on a 0.8% (w/v) low melting point agarose gel (Sigma) for 5 hrs at 30 V and 4°C. The marker lane was cut off, stained with ethidium bromide and photographed against a ruler on a UV transilluminator. From the sheared DNA lane, slices of gel were cut out that corresponded to the following size ranges: 0.4-0.75 Kb, 0.75-1.5 Kb, 1.5-2 Kb and 2-2.5 Kb. The DNA in each gel slice was cleaned up as follows. The gel was melted at 65°C and equilibrated at 42°C. Agar ACE enzyme (Promega) was added to give 7.5×10^{-3} Units/µl [final] and incubated for 30 min at 42°C to digest the agarose. The DNA from each slice was extracted once with 1:1 phenol:chloroform, once with chloroform and then ethanol precipitated (2.3.5). Each DNA size fraction was resuspended in 10 µl of T0.1E and 0.5 µl was run on a 0.8% gel (Fig.2-1B). The 0.75-1.5 Kb fraction was used to construct the library. To prepare the DNA for ligation it was subjected to a second end-repair step and concurrently 5' phosphorylated. The following components were combined in a 30 µl reaction: 8 µl of the 0.75-1.5 Kb DNA fraction, 1 × NEB buffer 2, 0.1 µg/µl BSA, 2.5 Units of Klenow (NEB), 1.5 Units of T4 polymerase (NEB), 5 Units of T4 polynucleotide kinase (NEB), 1 mM rATP (Promega) and 2.5 mM dNTPs. The reaction was incubated for 10 min at 37°C before being extracted once with 1:1 phenol:chloroform, once with chloroform and then ethanol precipitated. The DNA was finally resuspended in 10 µl of T0.1E and stored at -20°C.

2.5.3 Preparation of pGADT7 vector DNA

20 µg of pGADT7 plasmid was digested with SmaI in a 100 µl reaction: 1 × buffer J (Promega), 0.1 µg/µl BSA and 0.6 Units/µl of SmaI (Promega). The reaction was incubated for 3 hours at 25°C, 20 min at 65°C and cleaned up by ethanol precipitation. The cut plasmid was then dephosphorylated in a 100 µl reaction with: 1 × SAP buffer and 16 Units of SAP at 37°C for 30 min and 65°C for 20 min. The

SmaI & SAP treated vector was gel purified as described in the previous section and resuspended in 40 µl of 10 mM Tris pH 8.0.

2.5.4 Library construction

pGADT7 and genomic insert DNA were prepared as described in the previous two sections. The optimum ratio for vector to insert was determined empirically by setting up several ligations and testing a sample of transformants by colony PCR. Based on this, the preparative shotgun ligation was set up as a 30 µl reaction containing 1.5 µl SmaI & SAP treated pGADT7, 0.3 µl *M. thermautotrophicus* genomic insert DNA, 1 × Rapid ligation buffer and 0.3 Units/µl T4 DNA ligase. The ligation was incubated overnight at 16°C. 5 µl of ligation was transformed into each of three 100 µl aliquots of XL10-Gold ultracompetent cells (Stratagene) according to the manufacturer's instructions. The transformed cells were spread onto 18 SOB ampicillin (100 µg/ml) plates of 150 mm diameter. 3 plates were also spread with a 1 in 10 dilution of the transformed cells. After an overnight incubation at 37°C the transformants were recovered by resuspending the cells in liquid SOB ampicillin medium. For each plate, 8 ml of medium was added, the cells scraped into suspension, then pipetted off. Any residual cells were recovered with an extra 4 ml of medium and then added to the rest. The cell suspension from all 18 plates was pooled and centrifuged at 6,000 *g* for 15 min at 10°C. This yielded a 3 g pellet that was stored at -80°C until processing. The library DNA was recovered from the pellet with a plasmid maxi-prep kit (Qiagen), using 10 columns, giving a total yield of 370 µg in 2.7 ml.

2.5.5 Analysis of library size and coverage

After the library transformation, the 3 plates that were spread with a 1 in 10 dilution of the transformed XL10-Gold cells were analysed further. The mean colony count for these plates was 517. Therefore the total number of transformants was approximately 517 colonies × 10 × 18 plates = 93,060. In order to estimate the number of transformants that contained inserts, a sample of 100 colonies were picked and analysed by colony PCR (Fig.2-2). Colony PCR was performed using primers directed against sequences either side the pGADT7 MCS (pGADT7for and pGADT7rev). Of the 89 PCRs that worked, 81% showed inserts to be present. By comparing the PCR amplicons to the molecular mass markers, the mean insert size

was approximately 1 Kb. Therefore the size of the complete library was estimated to be: 93,060 colonies × 81% × 1 Kb = 75,379 Kb. This may be expressed as a 43 fold coverage of the *M. thermautotrophicus* genome (i.e. 1,751 Kb).

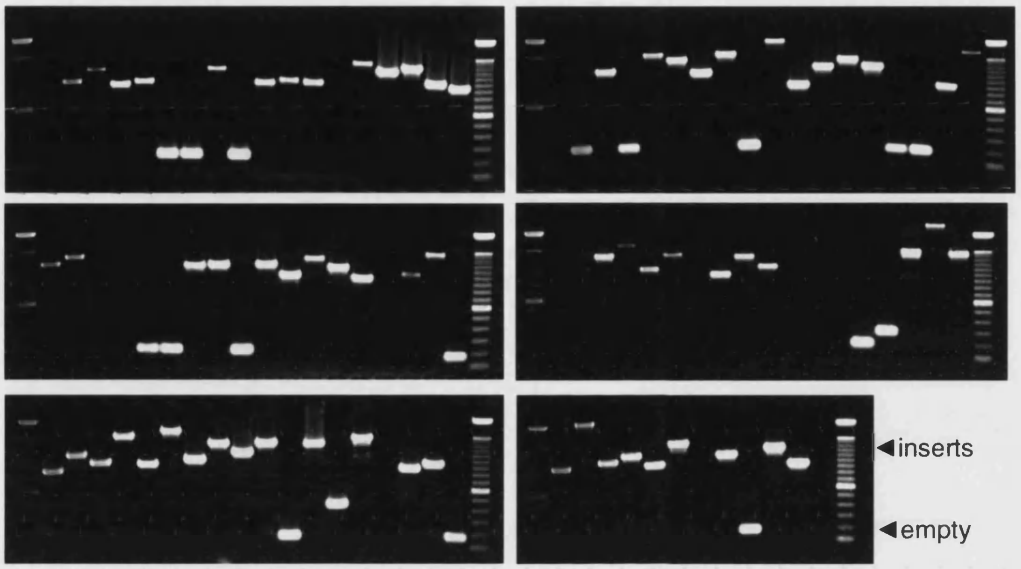


Figure 2-2 Colony PCRs of library clones. Of the 89 PCR amplicons 17 show empty vectors and 72 (81%) show clones with inserts (1.5% gels).

2.6 Screening Library against oriC, MCM and Cdc6-1

2.6.1 Screening the library against *oriC*

The library was screened for proteins that interact with the *M. thermautotrophicus oriC* region by using Clontech's Matchmaker One-hybrid system. The constructs pHisi-1/*oriC* and pHisi-1/*mth1888* were prepared as described in Section 2.4. These constructs were separately integrated into the genome of *S. cerevisiae* strain YM4271. The constructs were first linearised by digestion with AflII. The DNA was cleaned up by extraction with 1:1 phenol:chloroform, chloroform and then ethanol precipitation. The linearised constructs were transformed into strain YM4271. Non-linearised constructs were used as negative controls for integration. The transformed cells were plated onto SD/-His medium. During construct integration, the leaky *HIS3* expression from pHisi-1 allowed just enough colony growth for it to be used as a selectable marker. These strains were then stored as glycerol stocks. Integration of the pHisi-1/*oriC* construct was verified by colony PCR using OB13 and OB14 primers.

By analogy with *S. solfotaricus*, it was suspected that Cdc6-1 would interact with the *oriC*. This premise was used to test the system and determine the optimum conditions for screening. During integration, leaky *HIS3* expression was used as a selectable marker. When testing and screening for interactions, however, *HIS3* was used as a reporter gene. This was achieved by adding 3-amino-1, 2, 4-triazole (3AT) to the medium: a competitive inhibitor of His3p. The concentration of 3AT that was used needed to be high enough to suppress background growth due to leaky *HIS3* but low enough to allow growth due to interaction between *oriC* and Cdc6-1/AD fusion protein. The two reporter strains were transformed with pGADT7/*cdc6-1* and pGADT7 giving:

A) YM4271 pHisi-1/*oriC* pGADT7/*cdc6-1*

B) YM4271 pHisi-1/*oriC* pGADT7

C) YM4271 pHisi-1/*mth1888* pGADT7/*cdc6-1*

D) YM4271 pHisi-1/*mth1888* pGADT7

These were streaked onto SD/-His/-Leu plates containing 0, 15, 30, or 45 mM 3AT that were then incubated at 30°C for 5 days. 45 mM 3AT allowed A to grow but not B, C or D. It was decided that 45 mM 3AT should be used for library screening.

S. cerevisiae strain YM4271 pHisi-1/*oriC* was transformed with pGADT7 *M. thermautotrophicus* library DNA as described in Section 2.3.2. The main 60 ml culture was grown in YPAD medium instead of SD/-His since the pHisi-1/*oriC* construct had been integrated into the strain and also to improve transformation efficiency. Ten 360 µl transformations were performed, each using 1 µg of library DNA. The transformed cells were spread onto twenty 150 mm SD/-His/-Leu plates that were then incubated for 7 days at 30°C. A transformation efficiency of 5.97×10^5 CFU/µg/10^8 cells was achieved, giving a total of approximately 5.97×10^6 colonies. All the colonies were scraped into suspension, using 4 ml of sterile milliQ per plate, pooled, and then centrifuged for 5 min at 1,000 *g* and room temperature. The pellet was resuspended in an equal volume (to the pellet) of 65% (v/v) glycerol, 25 mM Tris-HCl, pH 8.0 and 0.1 M $MgSO_4$ before being stored in aliquots at -80°C. By plating dilutions of the stock onto SD/-His/-Leu the library titre was estimated to be 1.4×10^8 CFU/ml. 62.5 µl (~8.75×10^6 CFU) of library stock was then spread onto ten SD/-His/-Leu +45 mM 3AT selection plates and incubated for 5 days at

30°C. Only 50-100 colonies grew showing that 45 mM 3AT effectively suppressed background growth and that those colonies present probably contained *oriC* interactors.

45 positive colonies were picked and analysed by colony PCR with pGADT7for and pGADT7rev primers as described in Section 2.3.9. These amplicons were analysed further by restriction typing. Each PCR amplicon was digested directly with HaeIII in a 20 µl reaction containing: 10 µl of PCR amplicon, 1 × buffer C (Promega), 0.1 mg/ml BSA and 0.2 Units/µl HaeIII (Promega) then incubated for 1 hr at 37°C. The digestion products were run on a 1.5% (w/v) agarose gel. The 45 clones were sorted into 12 groups, each comprising those that had a similar banding pattern. It was assumed that clones in the same group contained redundant library clones, so one representative was chosen from each group for further analysis. pGADT7/library insert plasmid DNA was isolated from each representative colony by using ampicillin selection in *E. coli*. Each plasmid was transformed into the yeast negative control strain YM4271 pHisi-1/*mth1888*. Each of these transformants was streaked alongside the original positive strain onto a SD/-His/-Leu +45 mM 3AT plate. This was to establish whether each library/AD fusion protein would also interact with an unrelated fragment of *M. therautotrophicus* DNA (i.e. *mth1888*). If one did, it would suggest that the interaction of that fusion protein was not specific for *oriC*. 7 of 12 were *oriC* specific; they were submitted for sequencing.

2.6.2 Screening the Library against MCM and Cdc6-1

The library was screened for proteins that interacted with MCM and Cdc6-1 using Clontech's Matchmaker GAL4 Two-hybrid System 3. The library was screened for MCM and Cdc6-1 interactors using a similar strategy to that employed in the one-hybrid screening. The constructs pGBKT7/*mcm* and pGBKT7/*cdc6-1* were prepared as described in Section 2.4 and individually transformed into *S. cerevisiae* strain AH109 (2.3.2). These strains were cultured on SD/-Trp to maintain selection for the bait plasmid. Both of the strains were transformed with 10 µg of library plasmid, plated onto twenty 150 mm SD/-Trp/-Leu plates and incubated for 5 days at 30°C. The transformation efficiency for AH109 pGBKT7/*mcm* was 1.3×10^4 CFU/µg/10^8 cells and for AH109 pGBKT7/*cdc6-1* it was 2.3×10^4 CFU/µg/10^8 cells. So the total yield for each was in the order of 10^5 independant transformants. The colonies from

each transformation were scraped into suspension with sterile milliQ, pooled, and centrifuged at 1,000 *g* for 5 min at room temperature. Each pellet was resuspended in an equal volume (to the pellet) of 65% (v/v) glycerol, 25 mM Tris-HCl, pH 8.0 and 0.1 M $MgSO_4$ then stored in aliquots at -80°C. By plating dilutions onto SD/-Trp/-Leu the titres were determined as approximately 1.4×10^8 CFU/ml for the MCM library and 1.2×10^8 CFU/ml for the Cdc6-1 library. To screen each library, a volume equivalent to 5×10^6 CFU was diluted and spread evenly onto twenty SD/-Trp/-Leu/-His/-Ade plates to select for interactions. Each screen produced several hundred colonies on SD/-Trp/-Leu/-His/-Ade medium. This suggested that around 0.03% of the potential CFUs plated onto SD/-Trp/-Leu/-His/-Ade actually grew and therefore the selection was probably stringent enough. 56 colonies were picked from the MCM screen and 157 from the Cdc6-1 screen for analysis by colony PCR using pGADT7for and pGADT7rev primers (2.3.9). The PCR amplicons were further analysed by HaeIII restriction typing and sorted into groups of redundant library clones. pGADT7/library insert plasmid DNA was isolated from a representative colony from each group then transformed into AH109 together with empty pGBKT7 plasmid. These strains were streaked onto SD/-Trp/-Leu/-His/-Ade and any growth was taken to indicate a non-specific activation of the reporter genes. Those strains that activated the reporters specifically, i.e. only when the proper bait was present, were identified by sequencing of the library insert.

2.7 Preparing Expression Constructs

Novagen's pCDFDuet-1 co-expression vector was used to prepare expression constructs for the Cdc6-1 affinity co-purification experiments. *cdc6-1* was amplified by PCR (58°C / 1.5 min Phusion) with C1412startBamH1 and C1412stopHind3 primers. It was sub-cloned into MCS1 of pCDFDuet-1 with BamH1 and Hind3 restriction sites to make pCDFDuet-1/*cdc6-1*. *mth203* was amplified by PCR (57°C / 1.5 min Phusion) using C203startNde1 and C203endXho1 primers. *mth810* was amplified (60°C / 5 min Phusion) using Mth810startNde1 and C810endXho1 primers. *mcm* was amplified (57°C / 2.5 min Phusion) using Mth1770start and C1770endXho1 primers. Each of these three genes was cloned individually into MCS2 of the pCDFDuet-1/*cdc6-1* construct using Nde1 and Xho1 restriction sites.

- CDFDuet-1/*cdc6-1*/*mth203*
- CDFDuet-1/*cdc6-1*/*mth810*

- CDFDuet-1/*cdc6-1*/*mcm*

mth203 was also cloned into vector pET28a (Novagen) in order to produce His-tagged Mth203 protein in *E. coli*. The *mth203* gene was amplified by PCR (57°C / 1.5 min Phusion) with Mth203startNhe1 and Mth203stopHind3 primers. *mth203* was sub-cloned into pET28a using Nhe1 and Hind3 restriction sites.

- pET28a/*mth203*

2.8 Cdc6-1 Affinity Co-purification

An affinity co-purification (pull-down) experiment was performed to confirm the interaction between Cdc6-1 and Mth203 that had been observed with the yeast two-hybrid system. An interaction between Cdc6-1 and MCM was strongly suspected and therefore used as a positive control. An interaction between Cdc6-1 and Mth810 was not suspected and was used as a negative control. Three constructs were prepared in Novagen's CDFDuet-1 co-expression vector:

- CDFDuet-1/*cdc6-1*/*mth203*
- CDFDuet-1/*cdc6-1*/*mth810*
- CDFDuet-1/*cdc6-1*/*mcm*

In these constructs each pair of genes is co-expressed from two separate T7-lac promoters and ribosome binding sites. In all three constructs, Cdc6-1 was fused with an N-terminal His-tag and functioned as the bait in each pull-down. In each construct, a C-terminal S-tag was fused with a different prey protein (Mth203, Mth810 or MCM). The S-tag fusion sequence encodes a 15 residue peptide that binds with high affinity to the 104 residue S-protein derived from pancreatic ribonuclease A.

Protein expression, in *E. coli*, was directed by each of the constructs under identical conditions. The constructs were each transformed into the *E. coli* expression strain BL21 (DE3) Star Rosetta 2 and plated onto SOB plus 34 μg/ml chloramphenicol (to maintain pRARE2) and 50 μg/ml streptomycin (to maintain each CDFDuet-1 construct). Each strain was inoculated into 5ml LB chloramphenicol (34 μg/ml) streptomycin (50 μg/ml) in a sterile 50 ml centrifuge tube and incubated overnight at 30°C with shaking (200 rpm). 2 ml of overnight starter culture was inoculated into 400 ml LB chloramphenicol (34 μg/ml) streptomycin (50 μg/ml) in a sterile 1 L conical flask and incubated for 22 hours at 30°C with shaking (200 rpm). Cells were

harvested by centrifugation for 10 min at 10,000 *g* and 10°C and then stored at -20°C until use.

Protein extracts were prepared using a 1 g cell pellet from each of the three cultures. Each pellet was resuspended in 15 ml of lysis buffer (20 mM Tris-HCl, pH 7.9, 1 M NaCl, 10% [v/v] glycerol, 10 mM imidazole, 0.1 mM phenylmethanesulfonyl fluoride [PMSF], 5 mM β-mercaptoethanol and 25 Units/ml Benzonase nuclease [Novagen]). 35,000 Units of lysozyme (Sigma) was added to each of the suspensions and they were incubated for 20 min at 30°C with shaking. They were then sonicated on ice at full power 10 × 15 sec with intervals. The cell debris was pelleted for 15 min at 16,000 *g* and 10°C. A 100 µl sample of the clarified lysate was prepared for SDS-PAGE and stored at -20°C. The rest was kept on ice until it was applied to the His-bind column.

A 1 ml His-bind column was prepared for each of the 3 pull-downs. 2 ml of 50% (v/v) Ni-NTA agarose His-bind resin (Novagen) added to a gravity flow column. The resin was rinsed with 5 ml sterile milliQ, charged with 5 ml 50 mM $NiSO_4$ and equilibrated with 5 ml of binding buffer (20 mM Tris-HCl, pH 7.9, 1 M NaCl, 10% (v/v) glycerol and 10 mM imidazole). The columns were loaded with each clarified lysate and washed; first with 10 ml of binding buffer, then with 10 ml of wash buffer (20 mM Tris-HCl, pH 7.9, 1 M NaCl and 30 mM imidazole). Bound protein was eluted in 4 ml of elution buffer (20 mM Tris-HCl, pH 7.9, 1 M NaCl and 300 mM imidazole) and collected in 0.5 ml fractions. Samples were prepared for SDS-PAGE and stored at -20°C until being run on a 10% (w/v) gel (2.3.13).

Western blots were used to detect the bait and prey proteins in samples from the pull-downs. Duplicate western blots were prepared; one was probed for the His-tagged bait protein and the other for the three S-tagged prey proteins. First, samples from the three pull-downs were run on duplicate SDS-PAGE gels. The resolved proteins were then transferred to two polyvinylidene fluoride (PVDF) membranes by the standard semi-dry method described in Millipore's protein blotting handbook (4^{th} Ed.). Both membranes were then blocked with 2% (w/v) non-fat milk powder (Marvel) in TBS-T (10 mM Tris-HCl, pH 7.4, 0.9% [w/v] NaCl and 0.1% [v/v] Tween-20) for 1 hr at room temperature and then rinsed in TBS-T.

One of the blots was incubated overnight at 4°C with gentle rocking in TBS-T containing 1% (w/v) BSA and 0.05 μg/ml (1 in 4000) anti-His-tag monoclonal antibody (Novagen). It was rinsed 5 × with TBS-T followed by a 1 hr incubation at room temperature in horseradish peroxidase (HRP) conjugated anti-mouse IgG secondary antibody (Sigma) diluted 1 in 4000 in TBS-T with 1% (w/v) BSA. The blot was rinsed a further 5 × with TBS-T then treated with a peroxidase chemiluminescent substrate (GE Healthcare). The results were recorded using an EpiChemi 2 camera (Laboratory Products).

The second PVDF blot, after blocking and rinsing, was incubated for 1 hr at room temperature with HRP conjugated S-protein (Novagen) diluted 1 in 5000 with TBS-T 1% (w/v) BSA. It was washed 5 × with TBS-T, treated with chemiluminescent substrate and photographed as before.

2.9 Mth203 Expression and Purification

The *mth203* gene was cloned into pET28a to give a construct called pET28a/*mth203* (Section 2.7). The Mth203 protein was expressed as a fusion with an N-terminal His-tag. pET28a/*mth203* was transformed into *E. coli* BL21(DE3) pLysS and plated onto media containing 30 μg/ml kanamycin (to maintain the construct) and 34 μg/ml chloramphenicol (to maintain pLysS). The expression strain was inoculated into 5 ml of LB kanamycin (30 μg/ml) chloramphenicol (34 μg/ml) in a 50 ml centrifuge tube and incubated overnight at 30°C with shaking (200 rpm). A 1 L conical flask containing 400 ml of LB kanamycin (30 μg/ml) chloramphenicol (34 μg/ml) was inoculated with 2 ml of starter culture and incubated at 30°C with shaking (200 rpm) until an OD_{600} of 0.6-0.7 was reached. At this point the culture was induced with 1 mM IPTG [final] and incubated overnight at 30°C with shaking (200 rpm). The final OD_{600} reached 2.9. The cells were harvested by centrifugation for 10 min at 10,000 *g* and 10°C, giving ~2 g wet weight of cells per 400 ml culture, and stored at -20°C until use.

A 1 g aliquot of cells was resuspended in 15 ml of lysis buffer (20 mM Tris-HCl, pH 7.9, 1 M NaCl, 10% [v/v] glycerol, 10 mM imidazole, 0.1 mM PMSF and 5 mM β-mercaptoethanol). The resuspended cells were frozen with liquid nitrogen and thawed in a 30°C water bath before being incubated for 20 min at 30°C to allow the

released lysozyme (encoded by pLysS) to act. The suspension was sonicated on ice at full power for 10 × 15 sec with intervals. The debris was pelleted by centrifugation at 16,000 *g* for 15 min at 10°C. A sample of clarified lysate was prepared for SDS-PAGE and stored at -20°C. The clarified lysate was processed through a 1 ml His-bind column as described in Section 2.8. Samples of the 0.5 ml elution fractions were run on a 10% (w/v) SDS-PAGE gel. The third and forth fractions, which contained the most Mth203 protein, were pooled to give one sample for the next purification step.

Mth203 was further purified by size exclusion chromatography using a Superdex 200 16/26 column connected to an Äkta FPLC system (GE Healthcare). The column was first equilibrated with GF buffer (20 mM Tris-HCl, pH 7.9, 1 M NaCl and passed through a filter of pore size 0.22 μm). The column was then calibrated with molecular mass standards: blue dextran, β-amylase, alcohol dehydrogenase, BSA, carbonic anhydrase and cytochrome C. The Mth203 sample was passed through the column at a flow rate of 1 ml/min and collected in a series of 1 ml fractions. The best fractions, based on the A280 trace, were pooled to give a 7 ml sample. The protein was concentrated to 1.75 ml using a Vivaspin centrifugal filter (10 KDa cut-off). The concentrated Mth203 was aliquoted and stored at -20°C in buffer with the following final composition: 10 mM Tris-HCl, pH 7.9, 0.5 M NaCl, 50% (v/v) glycerol, 2 mM DTT (dithiothreitol) and 0.1 mM PMSF. The protein stock was analysed by Bradford assay and SDS-PAGE.

2.10 Mth203 ATPase assays

ATPase activity of the purified Mth203 protein was measured using a colorimetric assay based on the molybdate / malachite green reaction. The ATPase reaction composition was based on one developed for Has1p, a yeast DEAD box protein (Rocak *et al.*, 2005). A standard reaction contained: 20 mM HEPES acetate, pH 6.5, 50 mM K acetate, 2 mM $MgCl_2$, 2 mM DTT, 0.1 mg/ml BSA, 250 μM rATP and 360 nM Mth203 protein. This was incubated at 50°C using a PCR machine and samples were taken at 0.5 or 1 hour intervals. Alterations of these standard conditions and their effects are described in the results.

ATPase activity was determined by measuring the amount of inorganic phosphate (Pi) released. Inorganic phosphate forms a complex with molybdate and malachite green that absorbs at 630 nm. The malachite green reagent was prepared as follows and stored at 4°C: 0.03% (w/v) malachite green oxaolate, 8.3 mM Na molybdate, 0.7 M HCl and 0.05% (v/v) Triton X-100. An aliquot of this was diluted with 0.5 volumes of milliQ before each assay and kept at 4°C. When a time point was reached, a 50 µl sample of each reaction was transferred to a 96-well plate. 150 µl of the diluted malachite green reagent was added to each sample, mixed by pipetting and incubated for 20 min at room temperature. Each sample's A630 was then measured with a Tecan Spectra plate reader.

3 Origin Interaction Screen

3.1 Introduction

At present there is a wealth of evidence to suggest that archaeal Cdc6/Orc1 proteins are the archaeal origin recognition factors (Kasiviswanathan *et al.*, 2006, Robinson *et al.*, 2004, Shin *et al.*, 2003). However, at the time the work in this chapter was initiated evidence was sparse. The single Cdc6/Orc1 protein in *P. abysii* had been shown to associate specifically with the origin of replication in that organism by chromatin immunoprecipitation (Matsunaga *et al.*, 2001). However, it remained a possibility that Cdc6/Orc1 was recruited to the origin by another protein. Also, the question remained: in archaeal species with multiple Cdc6/Orc1 proteins (such as *M. thermautotrophicus* and *S. solfataricus*), which one/s interact with the origin? Is the situation in archaea like in bacteria where one protein (DnaA) binds to the origin cooperatively or like in eukaryotes where multiple different proteins interact with the origin as a complex (the ORC)? Are there additional factors in archaea that interact with the origin to regulate initiation of replication? For example, in bacteria, SeqA binds to newly replicated *oriC* sites and sequesters them from further rounds of initiation (Boye *et al.*, 1996). When the work that is presented in this chapter was undertaken, these questions remained relatively un-investigated. Therefore it was deemed worthwhile to try to identify exactly what proteins interact with the origin of replication in *M. thermautotrophicus*.

Several techniques were available for the identification / analysis of DNA-protein interactions. The yeast one-hybrid system was chosen as it offered several advantages over the alternatives. The DNA-binding proteins are expressed *in vivo,* meaning the assay is sensitive and binding conditions do not require optimisation. It is a powerful and convenient method for screening a library for potential interactions and the same library can also be used for two-hybrid screening.

However, using the yeast one-hybrid system to study *M. thermautotrophicus* DNA-protein interactions also had some potential limitations. For example the optimum temperature for *M. thermautotrophicus* growth is 65 °C and the one-hybrid assay was performed at 30 °C. In addition, the chromatin structure in *S. cerevisiae* is

different from the chromatin structure in *M. thermautotrophicus*. However, these potential limitations were accounted for by using controls and by optimising the system before screening.

The yeast one-hybrid system was developed and first used by Wang and Reed (1993) to clone the gene for a human transcription factor, OLF-1. The one-hybrid assay was derived from the two-hybrid concept (discussed further in Chapter 4) i.e. reconstitution of the DNA binding domain and activation domain of a transcriptional activator via the interaction of two proteins, the bait and prey. In the one-hybrid system, the bait protein is replaced with a DNA sequence element and the prey binds specifically to that sequence and activates a downstream reporter gene.

3.2 Aims and Rationale

The aim was to identify proteins that interact specifically with the origin of replication in *M. thermautotrophicus*. The first objective was to use a yeast one-hybrid system to screen a *M. thermautotrophicus* genomic library for genes that encode proteins that interact with the origin (Fig. 3-1). The second objective was to identify those positives from the screen that were specific for the origin *i.e.* they would not interact with an unrelated *M. thermautotrophicus* DNA sequence.

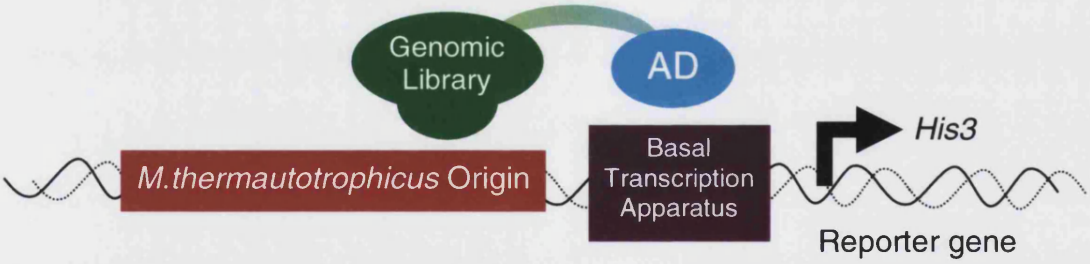


Figure 3-1 the yeast one-hybrid system was used to identify origin interacting proteins from a *M. thermautotrophicus* genomic library.

3.3 Results

3.3.1 Set-up of the Yeast One-hybrid System

A construct containing the *M. thermautotrophicus oriC* region and *HIS3* reporter gene was integrated into the genome of *S. cerevisiae* strain YM4271 (Section 2.6.1). Leaky *HIS3* expression from the reporter construct was used as a selectable marker

for integration. PCR was used to confirm that the construct had been successfully integrated. Genomic DNA was prepared from the reporter strain and used as a template in a PCR containing primers directed against *oriC* (Fig. 3-2). Since the reporter construct contained no origin for replication in yeast, and therefore could not be maintained as a plasmid, the result indicated that integration was successful.

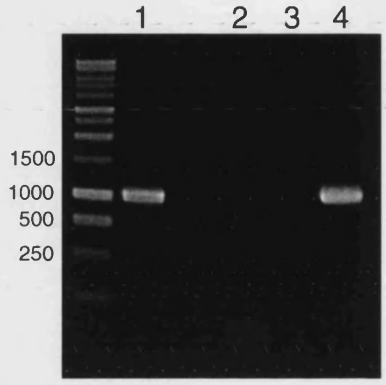


Figure 3-2 PCR confirming the integration of pHisi-1/*oriC* construct into *S. cerevisiae* strain YM4271. Yeast colony PCR using *oriC* specific primers. Lane: 1) YM4271 pHisi-1/*oriC*, 2) YM4271 only, 3) no template and 4) positive control.

During screening, background growth from leaky *HIS3* expression must be suppressed by adding 3AT to the medium. The optimum concentration of 3AT was determined by testing several concentrations for their ability to inhibit the growth of various negative controls but allow growth due to interaction between Cdc6-1 and the *oriC* (Fig. 3-3). One negative control was made by replacing the Cdc6-1/AD with the AD alone. The other negative control was made by replacing *oriC* with *mth1888*, an unrelated section of non-protein coding *M. thermautotrophicus* DNA. 45 mM 3AT (Fig. 3-3D) inhibited growth of the negative control strains but allowed growth of the strain containing the *oriC* and Cdc6-1. Therefore 45 mM 3AT was used during library interaction screening. A similar experiment demonstrated that MCM did not interact directly with the origin (Fig. 3-4).

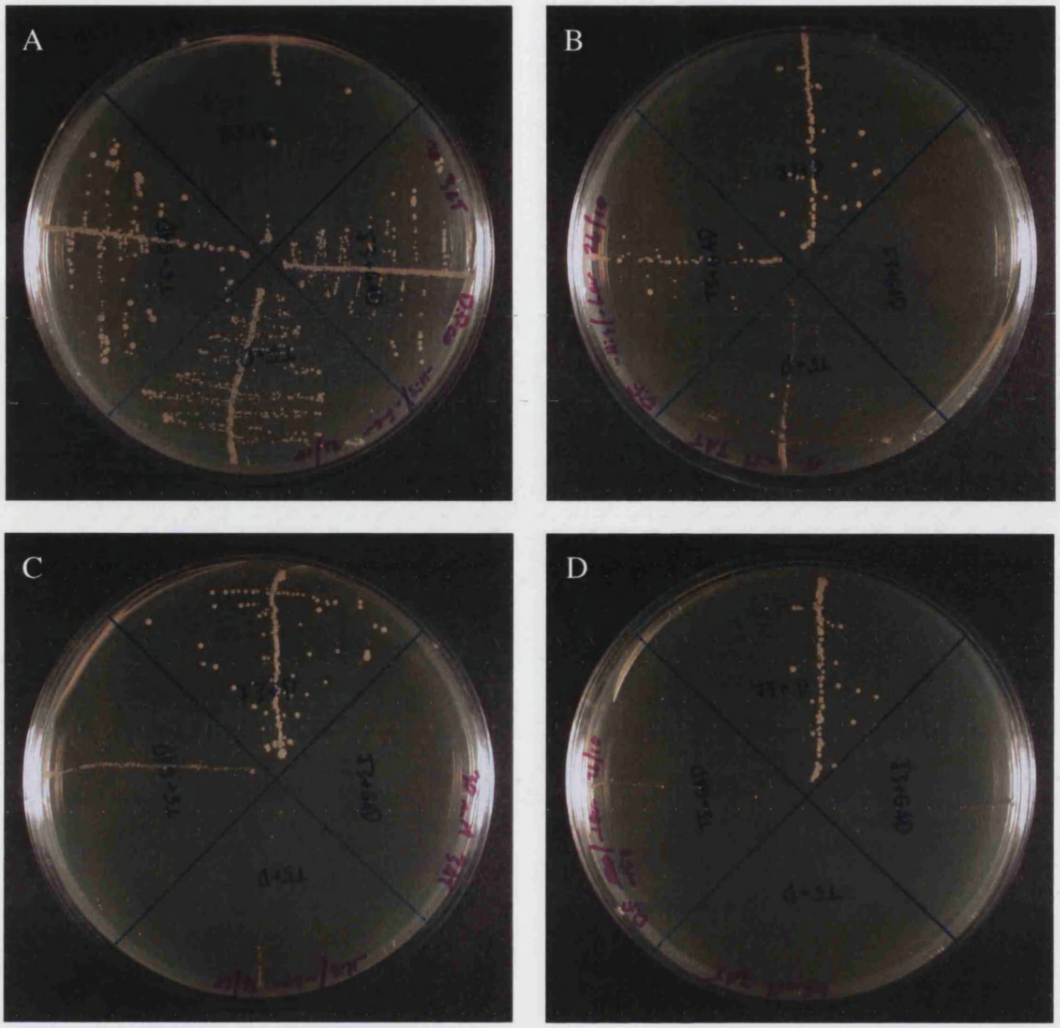


Figure 3-3 Optimisation of 3AT concentration in SD/–His/-Leu plates used to detect the interaction of Cdc6-1 with *oriC*. SD/–His/-Leu plates containing: A) no 3AT, B) 15 mM 3AT, C) 30 mM 3AT and D) 45 mM 3AT. Each plate was streaked with *S. cerevisiae* strain YM4271 containing: pHisi-1/*oriC* (integrated) + pGADT7/*cdc6-1* (top quadrant), pHisi-1/*oriC* (integrated) + pGADT7 (right), pHisi-1/*mth1888* (integrated) + pGADT7/*cdc6-1* (bottom) and pHisi-1/*mth1888* (integrated) + pGADT7 (left).

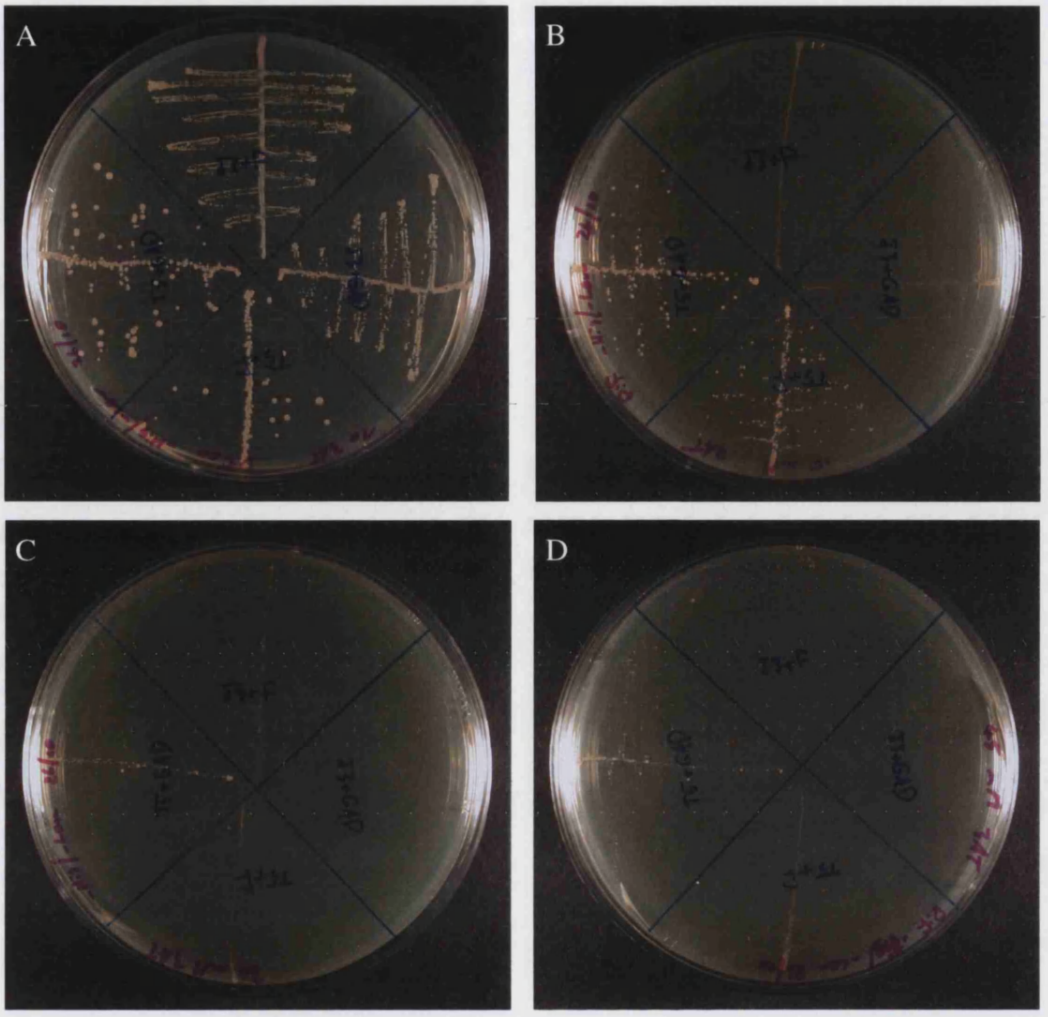


Figure 3-4 Interaction of MCM with *oriC*. SD/–His/-Leu plates containing: A) no 3AT, B) 15 mM 3AT, C) 30 mM 3AT and D) 45 mM 3AT. Each plate was streaked with *S. cerevisiae* YM4271 containing: pHisi-1/*oriC* (integrated) + pGADT7/*mcm* (top quadrant), pHisi-1/*oriC* (integrated) + pGADT7 (right), pHisi-1/*mth1888* (integrated) + pGADT7/*mcm* (bottom) and pHisi-1/*mth1888* (integrated) + pGADT7 (left).

3.3.2 Screening the Library

S. cerevisiae strain YM4271 pHisi-1/*oriC* was transformed with pGADT7 *M. thermautotrophicus* library giving approximately 5.97 × 10^6 transformants. These colonies were scraped into suspension and stored in a glycerol stock at a titre of approximately 1.4 × 10^8 CFU/ml. An amount of stock containing approximately 8.75 × 10^6 CFU was spread onto SD/–His/-Leu +45 mM 3AT interaction selection plates (Fig. 3-5).

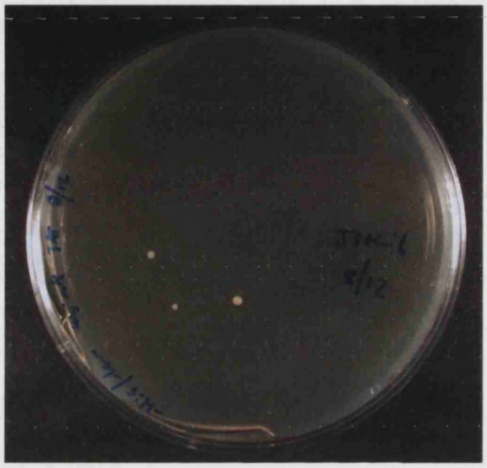


Figure 3-5 *S. cerevisiae* YM4271 pHisi-1/*oriC* (integrated) & pGADT7/library on SD/-His/-Leu/+45 mM 3AT selection medium. (One of ten such plates.)

Of a potential of 8.75 × 10^6 colonies only 50-100 grew, showing that 45 mM 3AT effectively suppressed background growth and that those colonies present probably contained *oriC* interactors. 42 colonies were picked and their inserts were characterised first by colony PCR (Fig. 3-6) and then by restriction typing (Fig. 3-7).

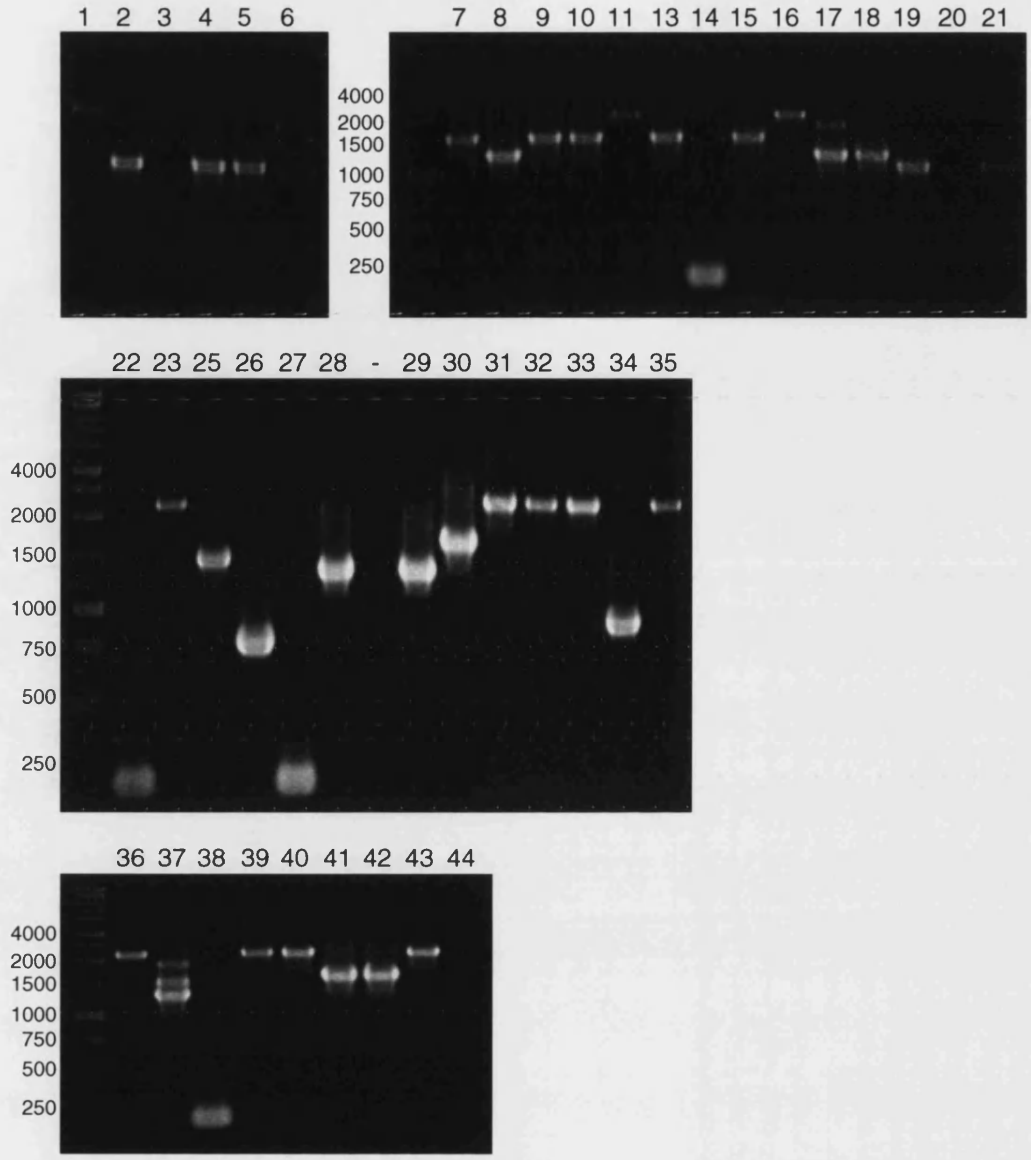


Figure 3-6 PCRs of *S. cerevisiae* YM4271 pHisi-1/*oriC* (integrated) & pGADT7/library colonies that tested positive for interaction. 42 colonies were picked from the *oriC* library interaction screen and analysed by colony PCR using pGADfor and pGADrev primers. Each PCR is labelled with the library clone identification number.

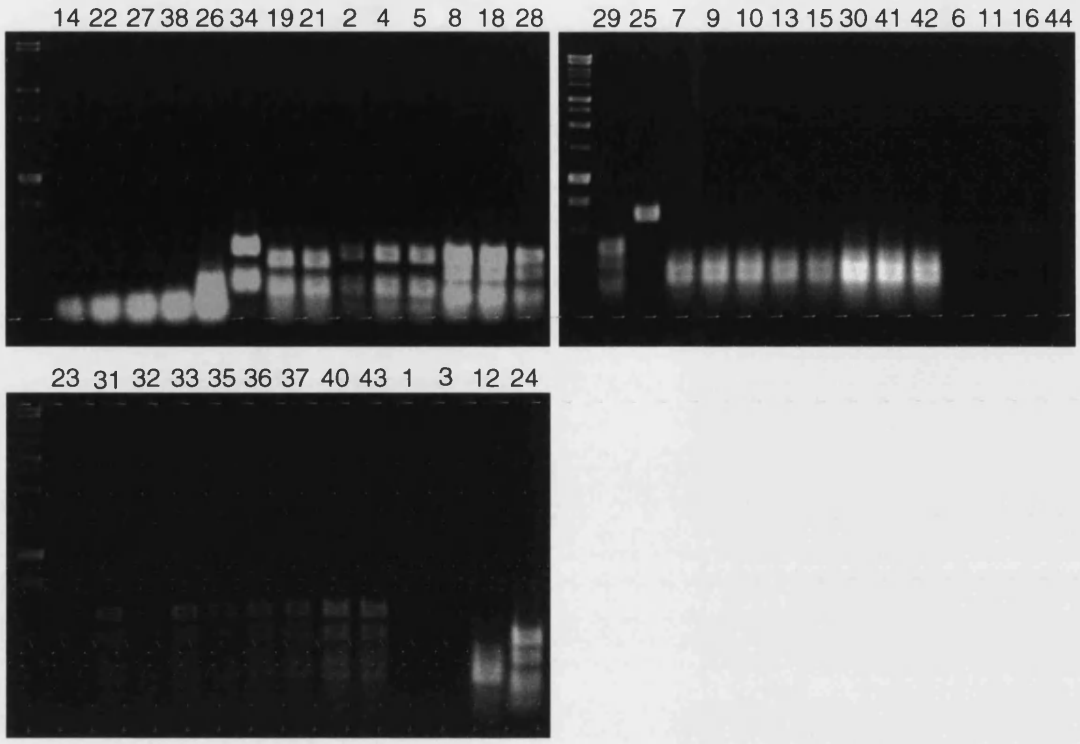


Figure 3-7 HaeIII digestion products of colony PCRs that tested positive for interaction. The colony PCR products were ordered according to size, digested with HaeIII and run on a gel. Each lane is labelled with the clone's identification number. Each clone was assigned to a group containing other clones with similar a banding pattern.

Based on these analyses, the 42 positives were sorted into 12 groups; each containing redundant library inserts (Table 3-1). One representative was chosen from each group for further analysis (Table 3-1).

Table 3-1 Groups of identical clones based on HaeIII digestion banding patterns.

Group	Representative
3	3
2, 4, 5	4
6	6
8, 18, 24, 28, 29	8
7, 9, 10, 12, 13, 15, 30, 41, 42	10
19, 21	19
14, 22, 27, 38	22
25	25
26	26
1, 11, 16, 23, 31, 32, 33, 35, 36, 37, 40, 43	31
34	34
44	44

3.3.3 Analysis of Positives

Non-specific DNA interactions were eliminated by testing each representative for interaction with *mth1888*, a negative control DNA element (Fig. 3-8). The 12 independent clones, identified during screening, were tested to see whether their interaction with *oriC* was due to a non-specific affinity for DNA. Each clone was transformed into a reporter strain containing a negative control bait DNA element (*mth1888*). These negative control strains were each tested for interaction alongside the original positive *oriC* strains (Fig. 3-8). 8 out of the 12 clones displayed a relatively specific interaction with *oriC* (4, 6, 19, 26 and 44). Those clones were then identified by DNA sequencing and comparison to the complete *M. thermautotrophicus* genome sequence (Table 3-2).

Of the 8 clones that were sequenced, 7 encoded fragments of the Cdc6-1 protein and 1 encoded part of a predicted glucose-1-phosphate thymidylyltransferase enzyme. The ranges of amino acid residues that were present in the Cdc6-1 library fragments are illustrated in Figure 3-9. This shows that all the Cdc6-1 fragments contained the WHD residues (287-383) as a minimum.

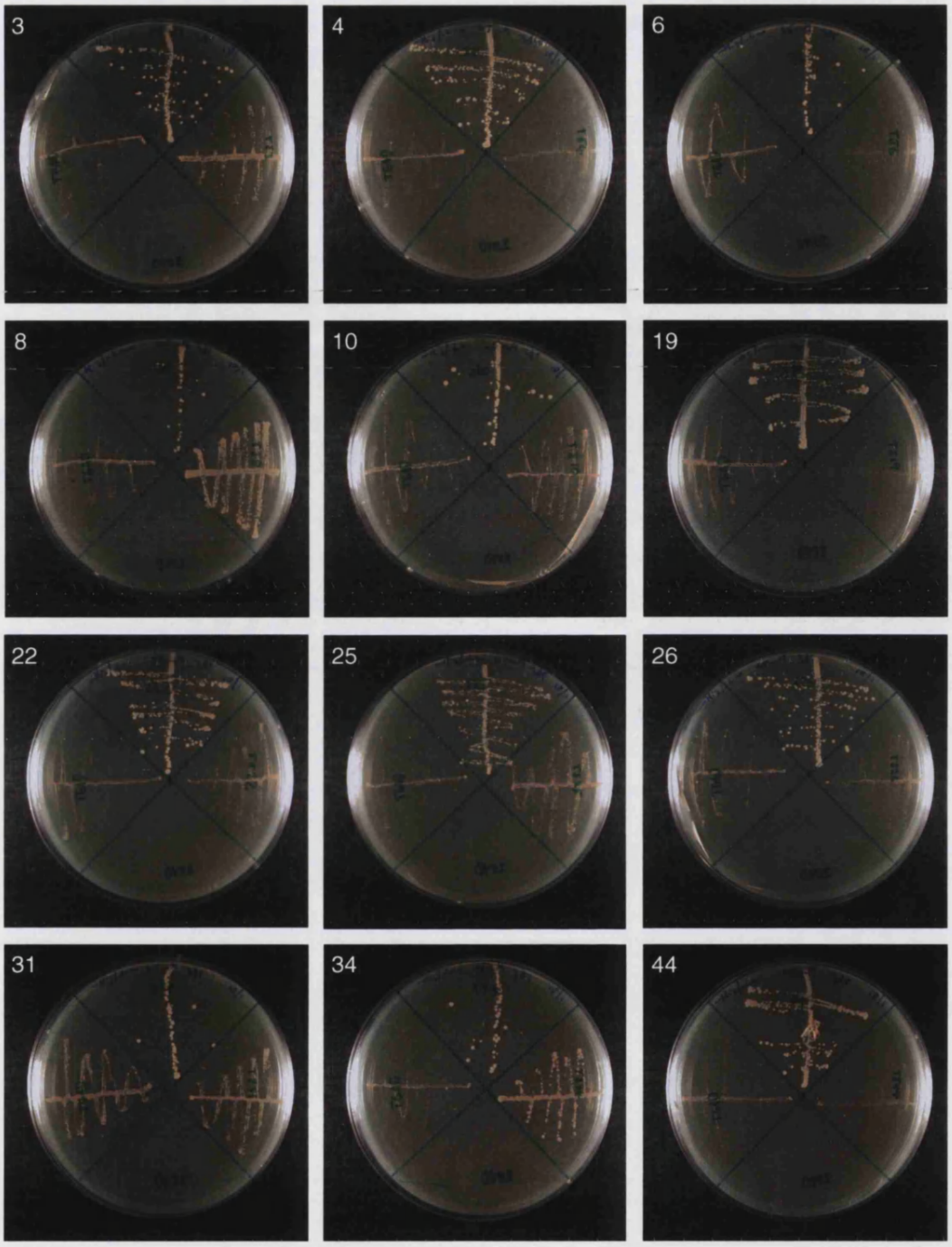


Figure 3-8 Comparison between interaction of positive library clones with either *oriC* or *mth1888* (a negative control). Growth in the right-hand quadrant indicated non-specific activation of the reporter genes. SD/-His/-Leu/+ 45 mM 3AT selection plates were streaked with *S. cerevisiae* YM4271 containing: pHisi-1/*oriC* (integrated) & pGADT7/library clone (indicated) (top quadrant), pHisi-1/*mth1888* (integrated) & pGADT7/library clone (right), pHisi-1/*oriC* (integrated) & pGADT7/empty (bottom) and pHisi-1/*mth1888* (integrated) & pGADT7/empty (left).

Table 3-2 Information about positive library clones that were specific for *oriC*.

ID of positive	Mth ORF code	Mth Genome Annotation*
3	Mth1412	Cdc6-1
4	Mth1412	Cdc6-1
6	Mth1412	Cdc6-1
10	Mth1412	Cdc6-1
19	Mth1412	Cdc6-1
26	Mth1589	glucose-1-phosphate thymidylyltransferase homolog
31	Mth1412	Cdc6-1
44	Mth1412	Cdc6-1

*Annotation taken from *M. thermautotrophicus* complete genome gi15678031 (Smith *et al.*, 1997). The Cdc6-1 isolates were independent clones.

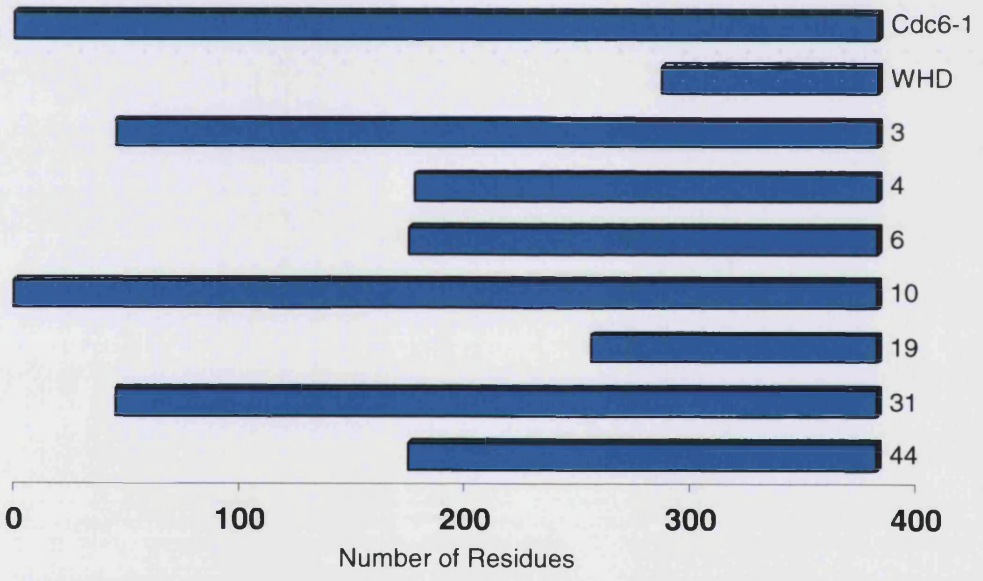


Figure 3-9 Sliding bar diagram showing the ranges of *oriC* interacting Cdc6-1 fragments. The extent of the full-length Cdc6-1 protein and its winged helix domain (WHD) are also shown.

3.4 Discussion

Cdc6-1 interacted specifically with *oriC*. This is consistent with results obtained in previous studies (Capaldi and Berger, 2004, Kasiviswanathan *et al.*, 2005). It is the first time that this observation has been made using the yeast one-hybrid system, however, as the previous studies employed a filter binding assay. This may attest to the validity of using the one-hybrid system to screen for *M. thermautotrophicus* DNA-protein interactions. Interactions have also been observed between the Cdc6-1 homologue and origin DNA in *P. abyssii* and *S. solfataricus* (Matsunaga *et al.*, 2001, Robinson *et al.*, 2004).

The Cdc6-1 library fragments that interacted with *oriC* all contained the WHD as a minimum. This suggests that the intact WHD is required for interaction, an observation that is consistent with previous findings in *M. thermautotrophicus* (Grabowski and Kelman, 2001, Capaldi and Berger, 2004, Kasiviswanathan *et al.*, 2005, Kasiviswanathan *et al.*, 2006) and *S. solfataricus* (De Felice *et al.*, 2004a).

Clone 19 is the shortest clone that interacts with *oriC* and contains only 31 additional residues to the WHD. A previous study suggested the WHD alone was not sufficient for interaction with DNA and that the AAA^+ domains may also be required (Kasiviswanathan *et al.*, 2005). Perhaps the additional 31 residues encoded by clone 19 are sufficient extra to enable the WHD to interact with *oriC*. Furthermore, the previous study used a random DNA substrate not *oriC*. Therefore, perhaps if an *oriC* DNA substrate had been used an interaction with the WHD may have been detected.

The yeast one-hybrid *oriC* screen also pulled out Mth1589, a gene that encodes a predicted glucose-1-phosphate thymidylyltransferase enzyme. However, it is probably a false positive due to fortuitous DNA binding by Mth1589. The *E. coli* homologue RffH catalyses one of the steps in the synthesis of enterobacterial common antigen, a cell surface glycolipid found in Gram-negative enteric bacteria (Sivaraman *et al.*, 2002). Assuming that the enzyme has a similar function in *M. thermautotrophicus* then it is hard to envisage a biologically rational explanation for its interaction with *oriC*.

MCM does not interact specifically with *oriC* (Fig. 3-4). This observation is consistent with data collected from an *in vivo* chromatin immunoprecipitation study in *P. abyssii* (Matsunaga *et al.*, 2001). It was suggested that *P. abyssii* MCM is recruited to the origin by Cdc6/Orc1 and then translocates along the genomic DNA with the replication fork. Cdc6-2 was not identified in the *M. thermautotrophicus* *oriC* interaction screen. This suggests that it does not interact specifically with *oriC*. It also corroborates a previous study that used a filter binding assay (Kasiviswanathan *et al.*, 2006). This observation is consistent with the hypothesis that *M. thermautotrophicus* Cdc6-1 acts as the origin recognition factor, whereas Cdc6-2 is the MCM helicase loader. However, it is in contrast to studies in *S. solfataricus* where all 3 Cdc6/Orc1 homologues (including a Cdc6-2 homologue) have been shown to interact with the replication origins.

Apart from Cdc6-1, no other proteins (except Mth1589) were identified in the *oriC* interaction library screen. This suggests that the *M. thermautotrophicus* genome encodes only one protein that interacts specifically with the *oriC* region, namely Cdc6-1. It is possible that other interactions do exist but were not identified during screening. One possibility is that the method or conditions used were not suitable for detecting those particular interactions. The observed Cdc6-1 *oriC* interaction, whilst not being a direct refutation, is at least support for the case against that premise. It would be unfeasible to test each and every *M. thermautotrophicus* protein individually under various conditions. Another possibility is that the library screen was not saturating. However, the fact that Cdc6-1 was pulled out many times suggests that the screen was very thorough and probably was saturating. Another caveat is that genes that encode toxic proteins will not be represented in the library; for instance any protein that might interfere with DNA replication or other central processes of the host cell.

The suggestion that Cdc6-1 appears to be the only true and specific *oriC* interactor is more evocative of bacterial DnaA than eukaryotic ORC. After all, Cdc6-1 is also structurally similar to DnaA and its gene (Mth1412) maps next to the *oriC* site in the *M. thermautotrophicus* genome. This suggestion would also mean that the only way for *M. thermautotrophicus* to recruit proteins specifically to the origin is via interaction with Cdc6-1.

4 MCM and Cdc6-1 Interaction Screens

4.1 Introduction

The MCM and Cdc6 proteins are key components of the eukaryotic DNA replication licensing and initiation mechanisms (Stillman, 2005, Robinson and Bell, 2005). There is evidence to suggest that similar (if simpler) mechanisms exist in archaea and that the archaeal MCM and Cdc6 homologues are directly involved in them. Through analogy with the eukaryotic system, it was hypothesised that the functions of archaeal MCM and Cdc6 are likely to involve physical interactions with other, as yet unidentified, protein components. Therefore, the aim of this chapter was to identify proteins that interact with MCM and Cdc6 in archaea.

Several techniques are available for the identification of protein interactions, including: yeast and bacterial two-hybrid systems, phage based expression cloning, co-immunoprecipitation, affinity co-purification and far western analysis (Ausubel, 1987, Sambrook and Russell, 2001). The yeast two-hybrid system was chosen for use in this project as it has several useful advantages. The bait and prey proteins are expressed and allowed to interact in an intracellular environment, which makes it a sensitive assay, capable of detecting relatively weak interactions. The mode of detection is based on the reconstitution of a transcriptional activator, which means the assay is capable of detecting transient interactions such as those between an enzyme and a protein substrate (e.g., protein kinases). It lends itself well to genomic level screening strategies. When positives are obtained from library screens, they are easily identified by sequencing of the prey plasmid. And finally, libraries constructed in a two-hybrid prey vector are also compatible with the one-hybrid system. However, using the yeast-two hybrid system to screen for *M. thermautotrophicus* protein interactions does have some potential limitations. Firstly, as with any two-hybrid screen, it is susceptible to false positives. This can be mitigated by checking positives for non-specific protein binding and by confirming specific interactions using an independent method. Secondly, interactions are tested under somewhat different conditions from those they would encounter *in vivo*. In this case: an assay temperature of 30 °C compared with 65 °C, aerobic environment compared with anaerobic, different protein folding machinery, etc. However, these

points are also true for the alternative methods and can be accounted for by using positive and negative controls. Furthermore, there are precedents for the yeast-two hybrid system being used successfully to study hyperthermophilic archaeal protein interactions, e.g., in *S. solfataricus* (Dionne and Bell, 2005) and in *M. thermautotrophicus* (Kasiviswanathan *et al.*, 2005). Two types of genetic library may be used in a two-hybrid screen: a cDNA library or a genomic DNA library. A genomic library was used in this project for several reasons. cDNA libraries are inherently biased towards highly transcribed genes, whereas all genes are equally represented in a genomic library. *M. thermautotrophicus* genes do not contain introns and only a small proportion of the *M. thermautotrophicus* genome is non-protein coding making it suitable for a genomic library.

The yeast two-hybrid principle is based in general on the modular nature of transcription factors and in particular on that of GAL4 (Fields and Song, 1989). GAL4 is a *S. cerevisiae* transcriptional activator that controls genes whose products are responsible for metabolising galactose. GAL4 has a DNA-binding domain (DNA-BD) that recognises a specific upstream activation sequence (UAS) and an activation domain (AD) that contacts the basal transcription apparatus to up-regulate transcription. A flexible tether connects the BD and AD to each other (Fig. 4-1).

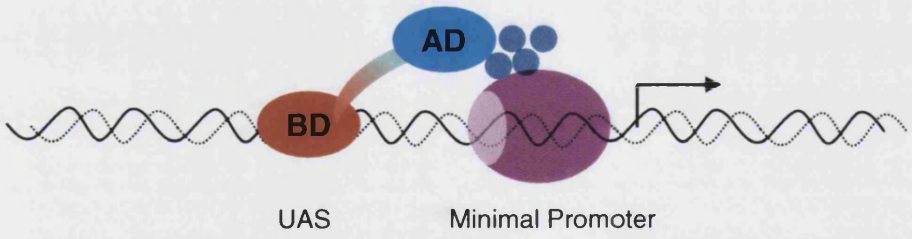


Figure 4-1 GAL4 Independent BD and AD (Adapted from Lewin, 2000)

The flexible tether allows the AD to contact the basal apparatus irrespective of the exact location and orientation of the BD. The two domains can function independently of each other. This allows each domain to function as part of a separate fusion protein. In a yeast two-hybrid assay the BD is fused with a bait protein and the AD is fused with a prey protein. If the bait and prey interact, the BD and AD are brought into proximity and will activate the transcription of any gene regulated by a GAL4 UAS (Fig. 4-2).

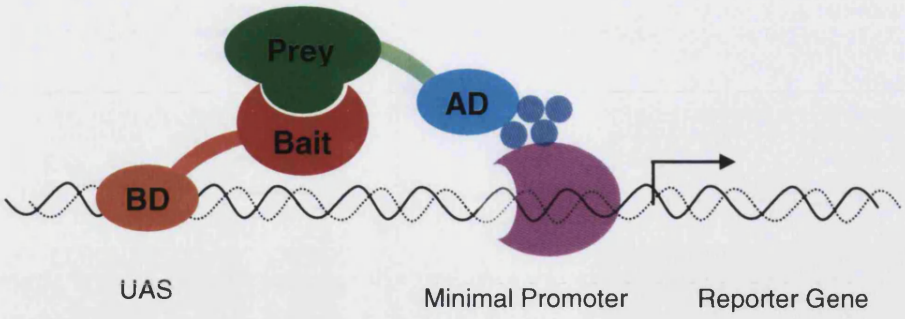


Figure 4-2 Model of Transcriptional Activation by Two-hybrid Proteins (Adapted from Fields and Song, 1989)

This system can be used to test the interaction between two known proteins or screen a library to find proteins that interact with a specific bait protein. Clontech developed the yeast two-hybrid system used in this project. In *S. cerevisiae* strain AH109 (Clontech), four reporter genes are regulated by GAL4 UASs (Fig. 4-3).

GAL1 UAS	GAL1 TATA	*HIS3*
GAL2 UAS	GAL2 TATA	*ADE2*
MEL1 UAS	MEL1 TATA	*lacZ*
MEL1 UAS	MEL1 TATA	*MEL1*

Figure 4-3 Reporter Constructs In Strain AH109. The *HIS3* and *ADE2* reporter genes enable nutritional selection for interaction by encoding enzymes of the histidine and adenine biosynthesis pathways, respectively. *lacZ* and *MEL1* enable colour selection by encoding β-galactosidase and α-galactosidase, respectively. GAL1, GAL2 and MEL1 are GAL4-responsive UAS and promoter elements.

Strain AH109 is normally auxotrophic for histidine and adenine but these mutations can be complemented by an interaction between GAL4 BD and AD. Strain AH109 also carries mutations in TRP1 and LEU2 making it auxotrophic for tryptophan and leucine. In this study the pGBKT7 and pGADT7 vectors (Clontech) were used; they encode the BD and AD fusion proteins, respectively. pGBKT7 and pGADT7 also carry a functional TRP1 or LEU2 gene, respectively. This enables nutritional selection for maintenance of these plasmids in strain AH109.

4.2 Aims and Rationale

The main aim of this chapter was to identify *M. thermautotrophicus* proteins that interact specifically with the initiation proteins MCM and Cdc6-1. The first objective was to validate the yeast two-hybrid system by using it to study some preliminary pair-wise control interactions. The next objective was to use the two-hybrid system to screen a *M. thermautotrophicus* genomic prey library for proteins that interact with MCM or Cdc6-1 baits. The last objective was to eliminate redundant positive library clones and then identify those that were specific for MCM or Cdc6-1 i.e. that would not interact with GAL4 BD alone.

4.3 Results

4.3.1 Preliminary Pair-wise Interactions

The three known *M. thermautotrophicus* DNA replication initiation protein genes (*mcm*, *cdc6-1* and *cdc6-2*) were cloned into the two-hybrid pGBKT7 bait and pGADT7 prey vectors. Various combinations of the three proteins were then tested for interaction with each other in order to assess the system and observe any interactions that had not been tested before. For each combination of proteins, the bait and prey plasmids were co-transformed into the *S. cerevisiae* AH109 reporter strain. Transformants were initially plated onto SD/-Trp/-Leu to select just for plasmid maintainance. They were then streaked onto selection medium and incubated for five days at 30°C before checking for growth. Two types of selection media were used: SD/-Trp/-Leu/-His low stringency (due to leaky HIS3 reporter gene) and SD/-Trp/-Leu/-His/-Ade high stringency (due to tightly regulated ADE2 reporter gene). The interactions were tested alongside negative controls made by replacing either the bait or prey plasmid with empty pGBKT7 or pGADT7, respectively. None of these negative control strains grew on high or low stringency media (Fig. 4-4). The results of the interaction tests are shown in Figure 4-4 and summarised in Table 4-1.

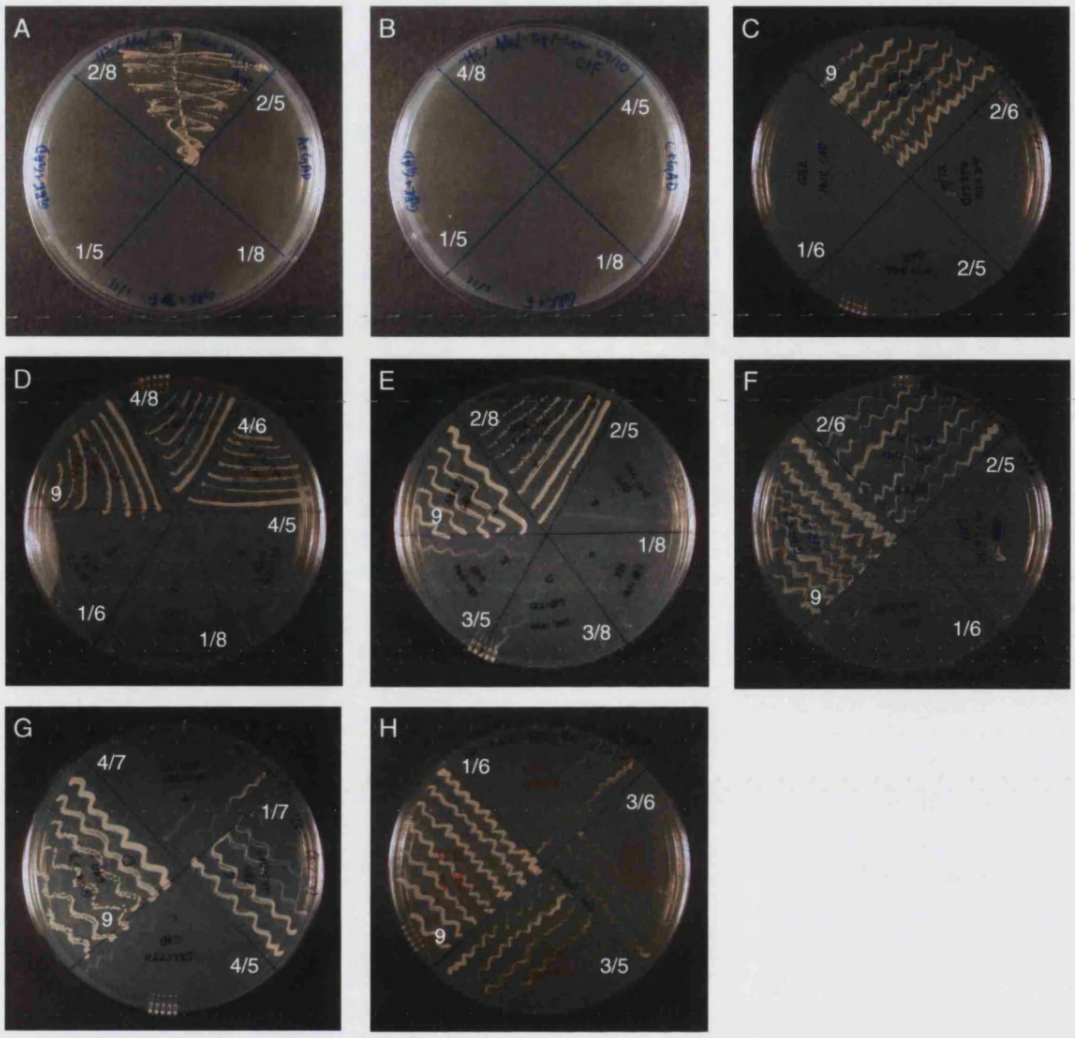


Figure 4-4 Interaction of various full-length proteins used to test the system. Plates A, B and C contained SD/-Trp/-Leu/-His/-Ade while plates D, E, F, G and H contained SD/-Trp/-Leu/-His. The plates were streaked with *S. cerevisiae* AH109 transformed with various combinations of bait and prey constructs (Table 4-1). The bait constructs are labelled: 1) pGBKT7, 2) pGBKT7/*cdc6-1*, 3) pGBKT7/*cdc6-2* and 4) pGBKT7/*mcm*. The prey constructs are labelled: 5) pGADT7, 6) pGADT7/*cdc6-1*, 7) pGADT7/*cdc6-2*, 8) pGADT7/*mcm* and 9) pGBKT7/53 pGADT7/T (positive control).

Table 4-1 Summary of results from Fig. 4-4

	pGBKT7/ *cdc6-1*	pGBKT7/ *cdc6-2*	pGBKT7/ *mcm*
pGADT7/*cdc6-1*	+/-	-/nt	+/nt
pGADT7/*cdc6-2*	nt/nt	nt/nt	-/nt
pGADT7/*mcm*	+ /+	-/nt	+/-

+) growth, -) no growth and nt) not tested. The left side of the slash mark represents growth on SD/-Trp/-Leu/-His while the right side represents growth on SD/-Trp/-Leu/-His/-Ade.

Cdc6-1 interacted with MCM as indicated by growth on the high stringency medium (Fig. 4-4A). Cdc6-1 also interacted with Cdc6-1, although this interaction was only indicated by growth on the low stringency medium (Fig. 4-4F). MCM interacted with MCM, although again this interaction only elicited growth on the low stringency medium (Fig. 4-4D). Cdc6-2 did not interact with either MCM or Cdc6-1 as no growth was observed even on the low stringency medium (Fig. 4-4E&H).

4.3.2 MCM Screen

The *M. thermautotrophicus* genomic library was screened to identify proteins that interacted with MCM. *S. cerevisiae* reporter strain AH109 was sequentially transformed (2.3.2), first with the pGBKT7/*mcm* bait construct and then with 10 μg of pGADT7/library DNA. The transformants were initially plated onto SD/-Trp/-Leu to select just for the maintenance of bait and prey plasmids. This generated ~1.3 × 10^5 individual transformants. The colonies were then scraped and pooled to give a glycerol stock with a titre of ~1.4 × 10^8 CFU/ml. A volume of this glycerol stock equivalent to 5 × 10^6 CFU was then diluted and spread evenly over 20 SD/-Trp/-Leu/-His/-Ade plates to select for interactions. Out of a potential of 5 × 10^6 colonies, only 1,500 grew (0.03%). This suggests that the selection conditions were stringent enough to allow only those transformants that produced MCM interacting prey proteins to grow.

56 of the positive colonies were picked and analysed by colony PCR using pGADT7for and pGADT7rev primers (Fig. 4-5). The sizes of the PCR amplicons

were estimated by agarose gel electrophoresis using marker size calibration curves. The 56 clones were then sorted into groups based on the PCR amplicon size (Table 4-2). By doing this, 3 separate groups were created. A single clone from each group was chosen to represent that group in further analysis. The predominant positive, found in 45 colonies, gave a ~600 bp long amplicon and was represented by clone 28.

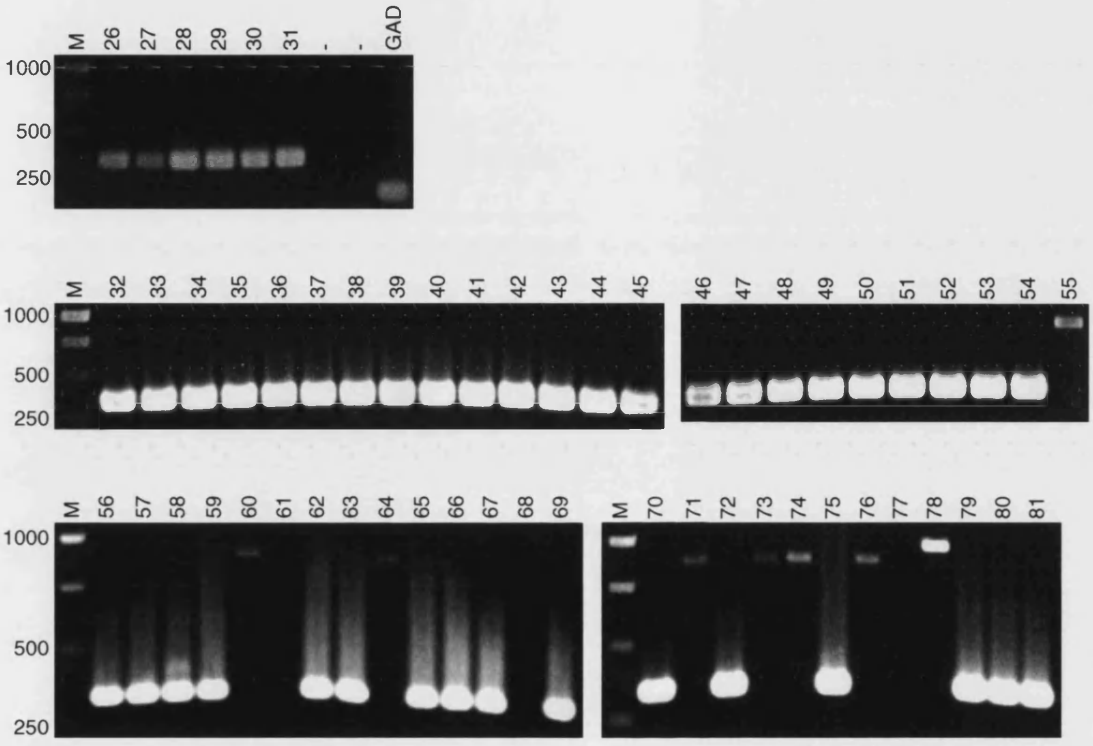


Figure 4-5 PCRs of a sample of AH109 pGBKT7/*mcm* pGADT7/library colonies that tested positive for interaction. 56 colonies were picked from the MCM / library interaction screen for analysis by PCR using pGADfor and pGADrev primers. GAD represents empty pGADT7 vector. PCR amplicons were analysed by 1.5% agarose gel electrophoresis. Marker sizes are indicated in base pairs.

Table 4-2 Groups of identical clones based on PCR amplicon sizes.

Group	Clones	Representative Clone
A	26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 62, 63, 65, 66, 67, 69, 70, 72, 75, 79, 80, 81	28
B	55, 60, 64, 71, 73, 74, 76	55
C	78	78

The next objective was to eliminate any false positives that arose due to non-specific activation of the reporter genes. The pGADT7/library clone DNA was isolated from each of the 3 representative positive yeast strains. Each clone was then transformed back into strain AH109 together with pGBKT7/empty plasmid. Those strains were then tested on SD/-Trp/-Leu/-His/-Ade plates alongside the original pGBKT7/*mcm* positives (Fig. 4-6). All three clones (28, 55 and 78) allowed strain AH109 to grow even when combined with the pGBKT7/empty plasmid. This suggests that the interaction was not specific for the MCM protein. In spite of that, the genes represented by clones 28 and 55 were sequenced and compared to the full genome sequence of *M. thermautotrophicus* (Table 4-3).

Figure 4-6 Comparison between interaction of positive library clones with either: pGBKT7/*mcm* or pGBKT7/empty (negative control). Growth in the right-hand quadrant indicated non-specific activation of the reporter genes. SD/-Trp/-Leu/-His/-Ade selection plates were streaked with *S. cerevisiae* AH109 containing the following: pGBKT7/*mcm* & pGADT7/library clone (indicated) (top quadrant), pGBKT7/empty & pGADT7/library clone (right), pGBKT7/*mcm* & pGADT7/empty (bottom) and pGBKT7/empty & pGADT7/empty (left).

Table 4-3 Information about positive library clones for MCM.

ID of Positive	Mth ORF Code	Mth Genome Annotation*
28	886	Conserved hypothetical protein
55	1074	Putative membrane protein

*Annotation taken from *M. thermautotrophicus* complete genome gi15678031 (Smith *et al.*, 1997)

4.3.3 Cdc6-1 Screen

The *M. thermautotrophicus* genomic library was screened to identify proteins that interacted with Cdc6-1. *S. cerevisiae* reporter strain AH109 was sequentially transformed, first with the pGBKT7/*cdc6-1* bait construct and then with 10 μg of pGADT7/library DNA. The transformants were initially plated onto SD/-Trp/-Leu to select for just the maintenance of bait and prey plasmids. This generated ~2.3×10^5 individual transformants. The colonies were then scraped and pooled to give a glycerol stock with a titre of ~1.2×10^8 CFU/ml. A volume of this glycerol stock equivalent to 5×10^6 CFU was then diluted and spread evenly over 20 SD/-Trp/-Leu/-His/-Ade plates to select for interactions. Out of a potential of 5×10^6 colonies, only 1,520 grew (0.03%). This suggests that the selection conditions were stringent enough to allow only those transformants that produced Cdc6-1 interacting prey proteins to grow.

157 of the positive colonies were picked and analysed by colony PCR using pGADT7for and pGADT7rev primers (Fig. 4-7 next page). The sizes of the PCR amplicons were estimated by agarose gel electrophoresis using marker size calibration curves. The amplicons were then ordered according to size, digested with HaeIII restriction enzyme and run on a 1.5% agarose gel (Fig. 4-8). The 157 library clones could then be sorted, according to their banding patterns, into groups of identical clones. By doing this, 40 groups were created. A single clone from each group was chosen to represent that group in further analysis (Table 4-4).

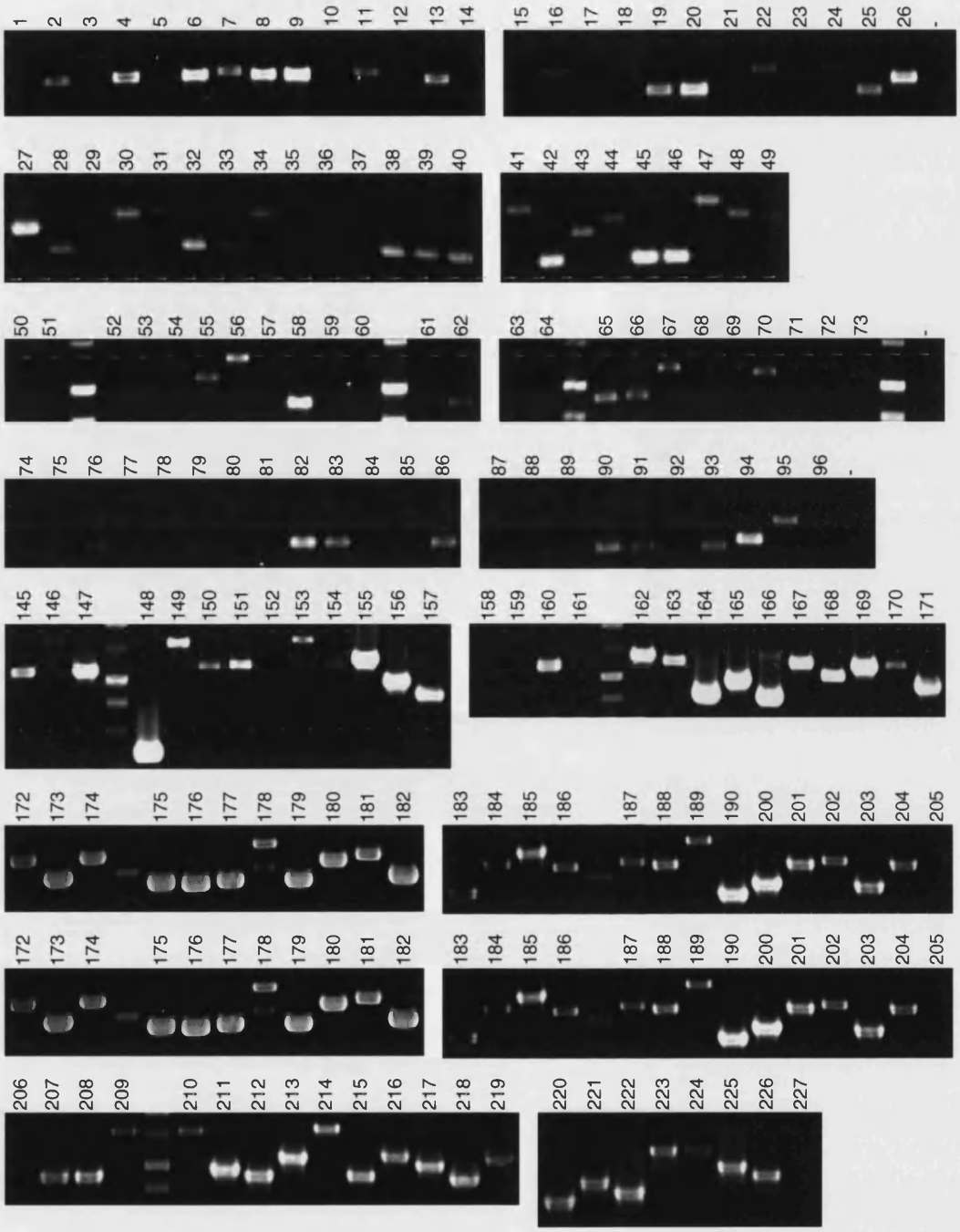


Figure 4-7 PCRs of a sample of *S. cerevisiae* AH109 pGBKT7/*cdc6-1* & pGADT7/library colonies that tested positive for interaction. 157 colonies were picked from the Cdc6-1 library interaction screen and analysed by colony PCR using pGADfor and pGADrev primers. Each PCR is labelled with the library clone identification number. 1.5% agarose gel electrophoresis.

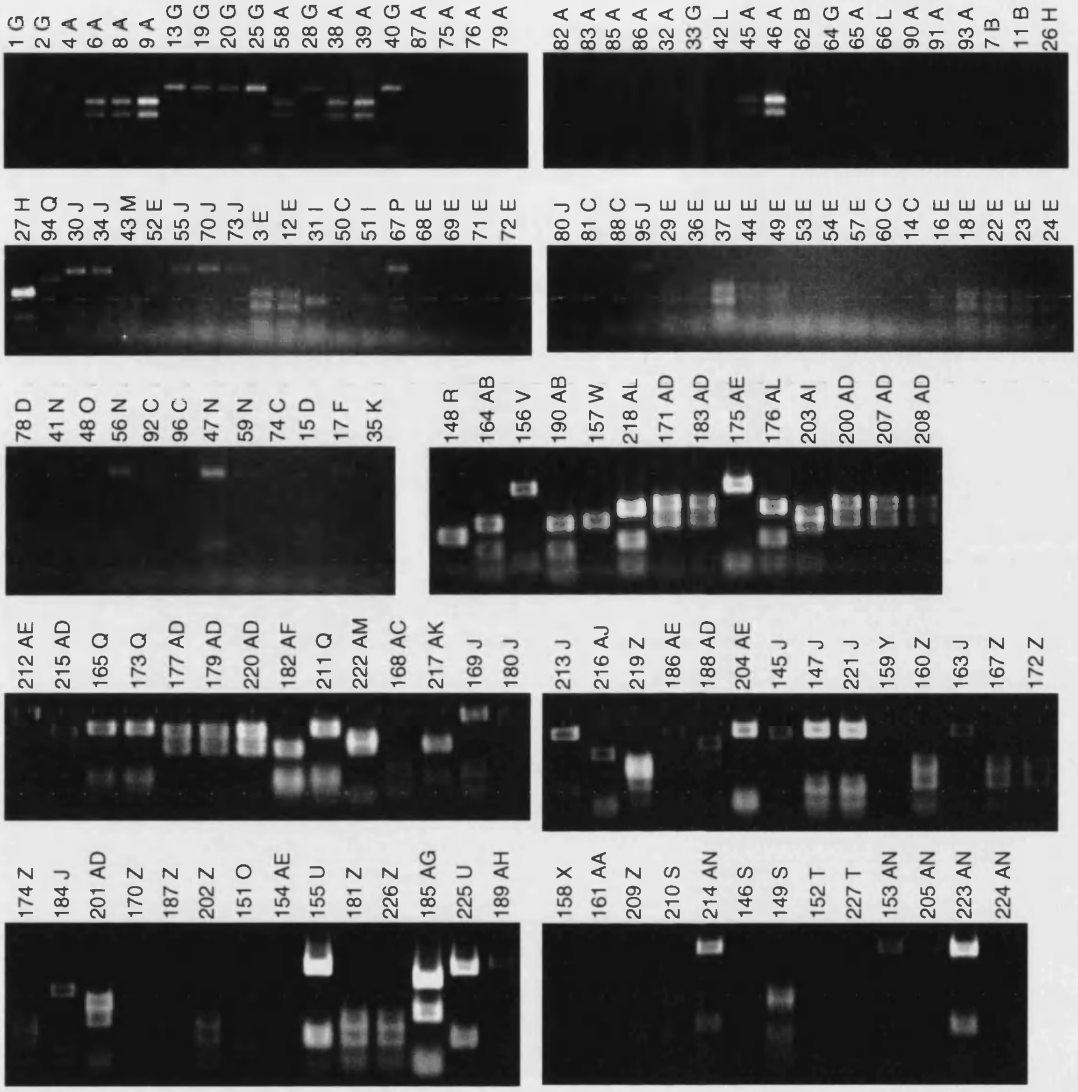


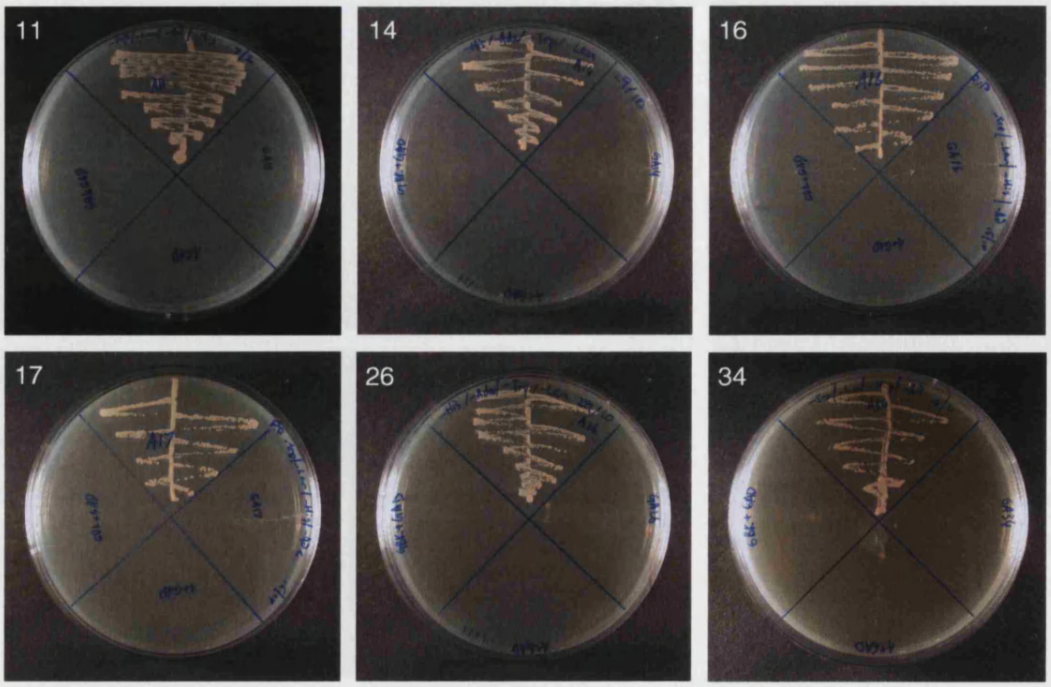
Figure 4-8 HaeIII digestion products of colony PCRs that tested positive for interaction. The colony PCR amplicons were ordered according to size, digested with HaeIII and run on a 1.5% agarose gel. Each lane is labelled with the clone's identification number. Each clone was assigned to a group containing other clones with a similar banding pattern. The group names are given as the letters written after each clone number.

Table 4-4 40 groups of identical clones based on HaeIII digestion banding patterns.

Group name	Clones in group	Representative clone	Insert size approx. (bp)
A	4, 6, 8, 9, 32, 38, 39, 45, 46, 58, 65, 75, 76, 79, 82, 83, 85, 86, 87, 90, 91, 93	4	900
B	7, 11, 62	11	1000
C	14, 50, 60, 74, 81, 88, 92, 96	14	1280
D	15, 78	15	1500
E	3, 12, 16, 18, 22, 23, 24, 29, 36, 37, 44, 49, 52, 53, 54, 57, 68, 69, 71, 72	16	1280
F	17	17	1580
G	1, 2, 13, 19, 20, 25, 28, 33, 40, 64	20	900
H	26, 27	26	1060
I	31, 51	31	1220
J	30, 34, 55, 70, 73, 80, 95, 145, 147, 163, 169, 180, 184, 213, 221	34	1180
K	35	35	1700
L	42, 66	42	940
M	43	43	1150
N	41, 47, 56, 59	47	1420
O	48, 151	48	1320
P	67	67	1240
Q	94, 165, 173, 211	94	1020
R	148	148	290
S	146, 149, 210	149	1620
T	152, 227	152	1620
U	155, 225	155	1220
V	156	156	740
W	157	157	760
X	158	158	1580
Y	159	159	1140
Z	160, 167, 170, 172, 174, 181, 187, 202, 209, 219, 226	160	1140
AA	161	161	1580
AB	164, 190	164	720
AC	168	168	1000
AD	171, 177, 179, 183, 188, 200, 201, 207, 208, 215, 220	171	820
AE	154, 175, 186, 204, 212	175	830
AF	182	182	930
AG	185	185	1280

AH	189	189	1560
AI	203	203	860
AJ	216	216	1100
AK	217	217	1000
AL	176, 218	218	800
AM	222	222	980
AN	153, 205, 214, 223, 224	223	1700

The next objective was to eliminate any false positives that were due to non-specific activation of the reporter genes. Therefore, the pGADT7/library plasmid DNA was isolated from each of the 40 representative positive yeast strains. Each clone was then transformed back into strain AH109 together with pGBKT7/empty plasmid. Those strains were then tested on SD/-Trp/-Leu/-His/-Ade plates alongside the original pGBKT7/*cdc6-1* positives. Based on the results of this test, the clones were divided into two groups containing: 14 clones that displayed specific (Cdc6-1 dependent) interaction (Fig. 4-9) and 26 that displayed non-specific (Cdc6-1 independent) interaction (Fig. 4-10). The 14 clones that showed specific interaction with Cdc6-1 were then identified by DNA sequencing and comparison to the complete *M. thermautotrophicus* genome sequence (Table 4-5).



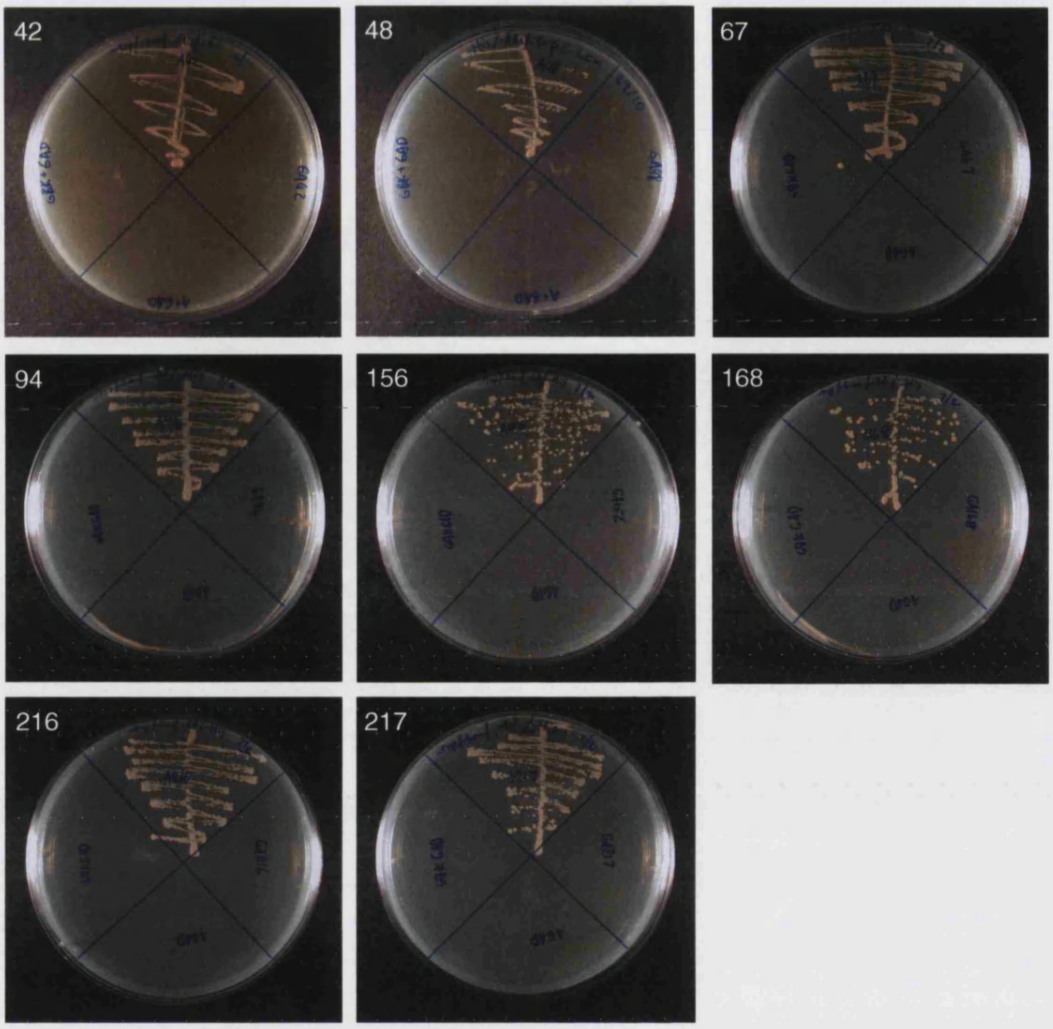
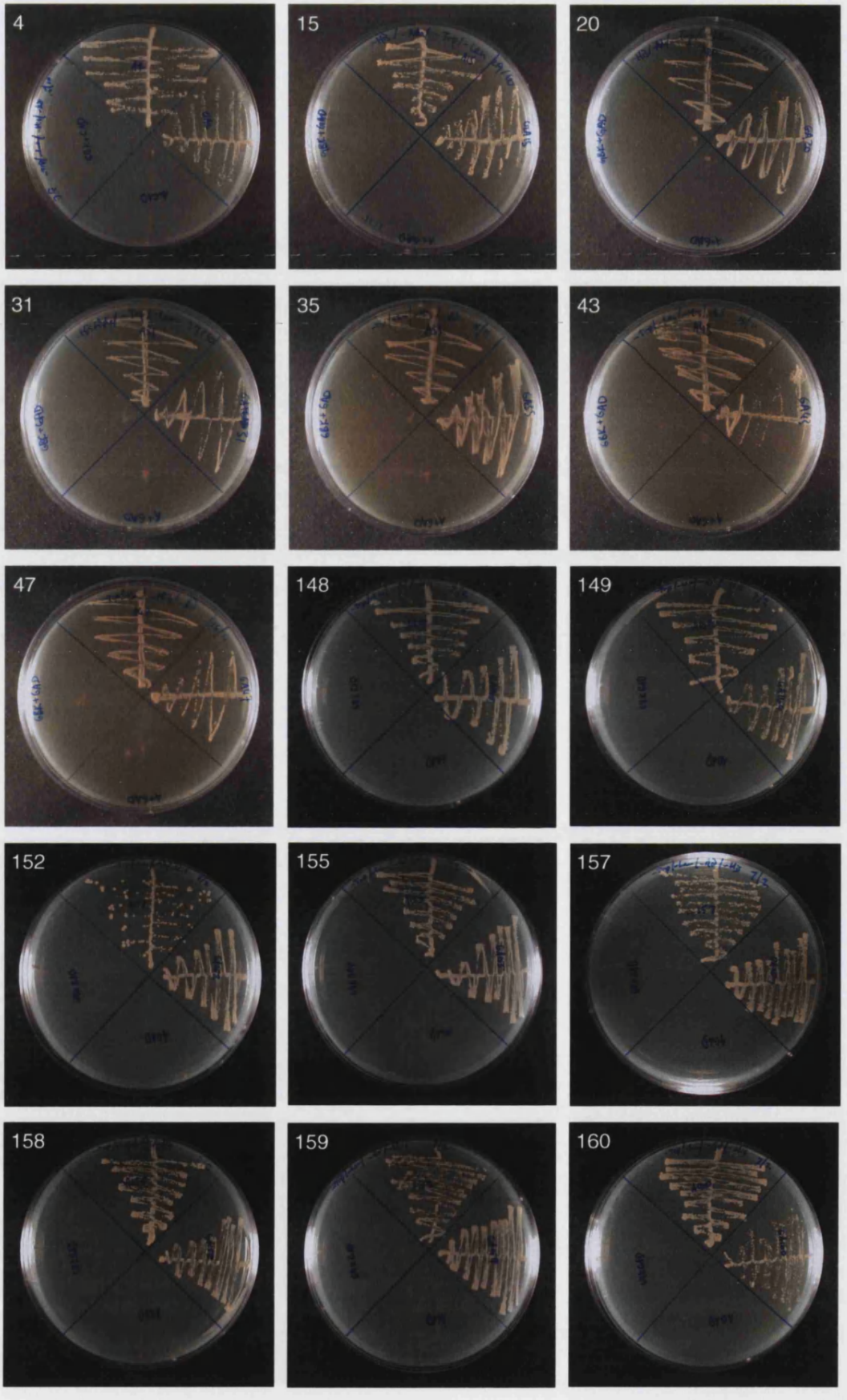


Figure 4-9 Comparison between interaction of positive library clones with either: pGBKT7/*cdc6-1* or pGBKT7/empty (negative control); clones that were SPECIFIC. Growth in the right-hand quadrant would have indicated non-specific activation of the reporter genes. SD/-Trp/-Leu/-His/-Ade selection plates were streaked with *S. cerevisiae* AH109 containing the following: pGBKT7/*cdc6-1* & pGADT7/library clone (indicated) (top quadrant), pGBKT7/empty & pGADT7/library clone (right), pGBKT7/*cdc6-1* & pGADT7/empty (bottom) and pGBKT7/empty & pGADT7/empty (left).


4
15
20
31
35
43
47
148
149
152
155
157
158
159
160

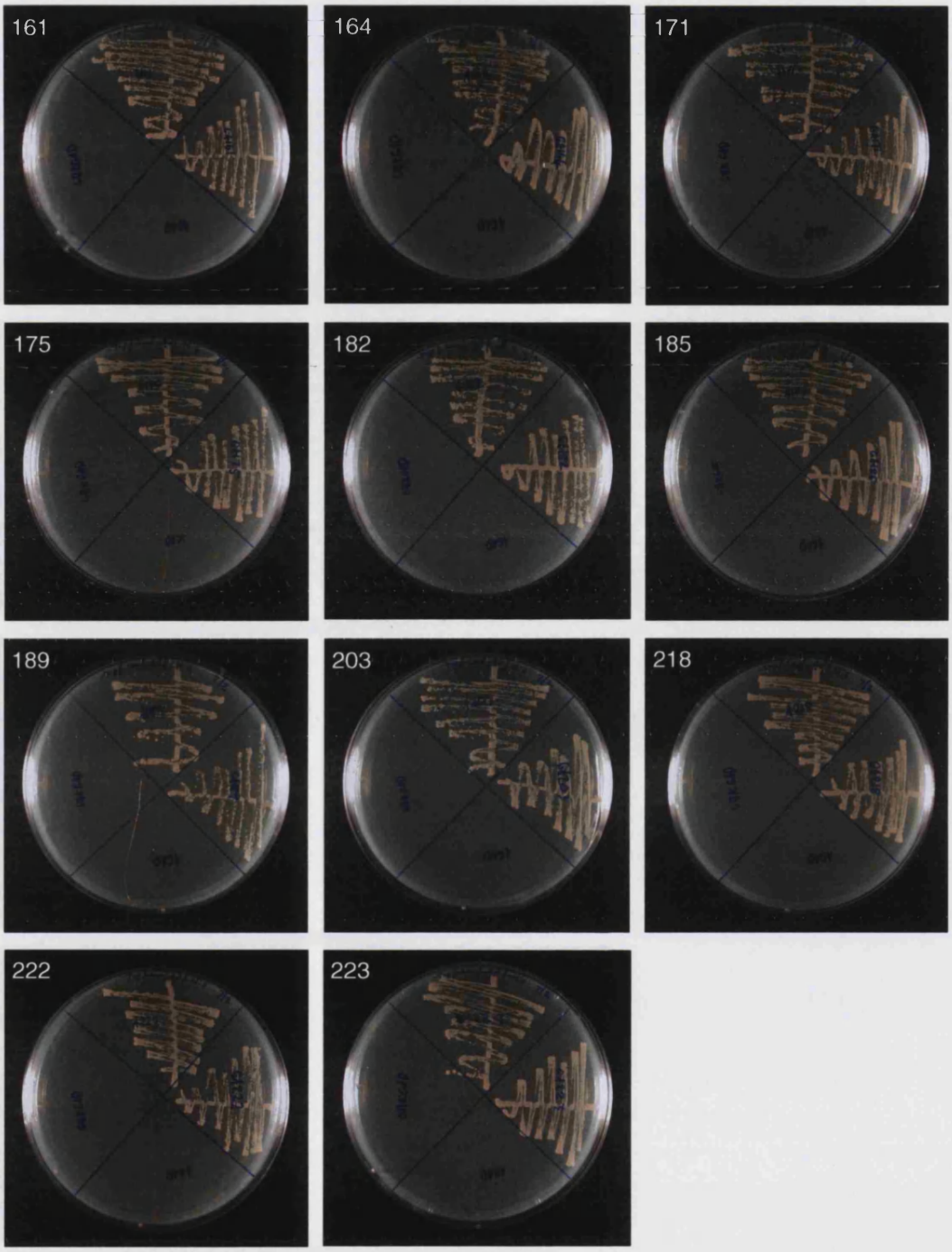


Figure 4-10 Comparison between interaction of positive library clones with either: pGBKT7/*cdc6-1* or pGBKT7/empty (negative control); clones that were NON-SPECIFIC. Growth in the right-hand quadrant indicated non-specific activation of the reporter genes. SD/-Trp/-Leu/-His/-Ade selection plates were streaked with *S. cerevisiae* AH109 containing the following: pGBKT7/*cdc6-1* & pGADT7/library clone (indicated) (top quadrant), pGBKT7/empty & pGADT7/library clone (right), pGBKT7/*cdc6-1* & pGADT7/empty (bottom) and pGBKT7/empty & pGADT7/empty (left).

Table 4-5 Information on positive library clones that were specific for Cdc6-1.

ID of Positive	Mth ORF Code	Mth Genome Annotation*
A11	203	ATP-dependent RNA helicase, eIF-4A family
A14	203	ATP-dependent RNA helicase, eIF-4A family
A16	1408	cobalamin biosynthesis protein G
A17	446	sensory transduction regulatory protein
A26	1907	conserved hypothetical protein
A34	151	methyl coenzyme M reductase system, A2 homolog
A42	492	ATP-dependent RNA helicase homolog
A48	1412	Cdc6-1
A67	1770	MCM
A94	1624	DNA topoisomerase I
A156	1458	hypothetical protein
A168	1215	fibrillarin-like pre-rRNA processing protein
A216	1715	phycocyanin alpha phycocyanobilin lyase CpcE related
A217	656	ATP-dependent RNA helicase related protein

*Annotation taken from *M. thermautotrophicus* complete genome gi15678031 (Smith *et al.*, 1997)

Table 4-5 outlines 14 clones that were identified by the screen for proteins that interacted specifically with Cdc6-1. Two of the genes identified were Mth1770, which encodes the MCM helicase protein and Mth1412, which encodes Cdc6-1 itself. These results are consistent with the findings of Section 4.3 in which full-length MCM and Cdc6-1 prey proteins were shown to interact with the Cdc6-1 bait. Another striking observation from Table 4-5 is that three of the genes identified are annotated as "ATP-dependent RNA helicases" (Mth203, Mth492 and Mth656). Therefore, these three proteins were picked for further investigation. Mth1907 was also picked as its annotation is "conserved hypothetical protein".

The next objective was to test full-length Mth203, Mth492, Mth656 and Mth1907 prey proteins for interaction with Cdc6-1. It was important to do this because the library clones all encoded only fragments of the respective *M. thermautotrophicus* proteins as the prey library was constructed by randomly shearing genomic DNA, thus ensuring an even coverage of the *M. thermautotrophicus* genome. As full-length proteins, only Mth203 interacted with Cdc6-1, whereas Mth492, Mth656 and

Mth1907 did not (Fig. 4-11 and Table 4-6). Mth810 has a degree of sequence similarity to Mth203, Mth492 and Mth656 but was not identified in the screen. It was therefore used as an additional negative control to test whether Cdc6-1 had a general tendency to interact with that type of protein. Mth810 did not interact with Cdc6-1 (Fig. 4-11). The results of the full-length protein interaction tests are summarised in Table 4-6.

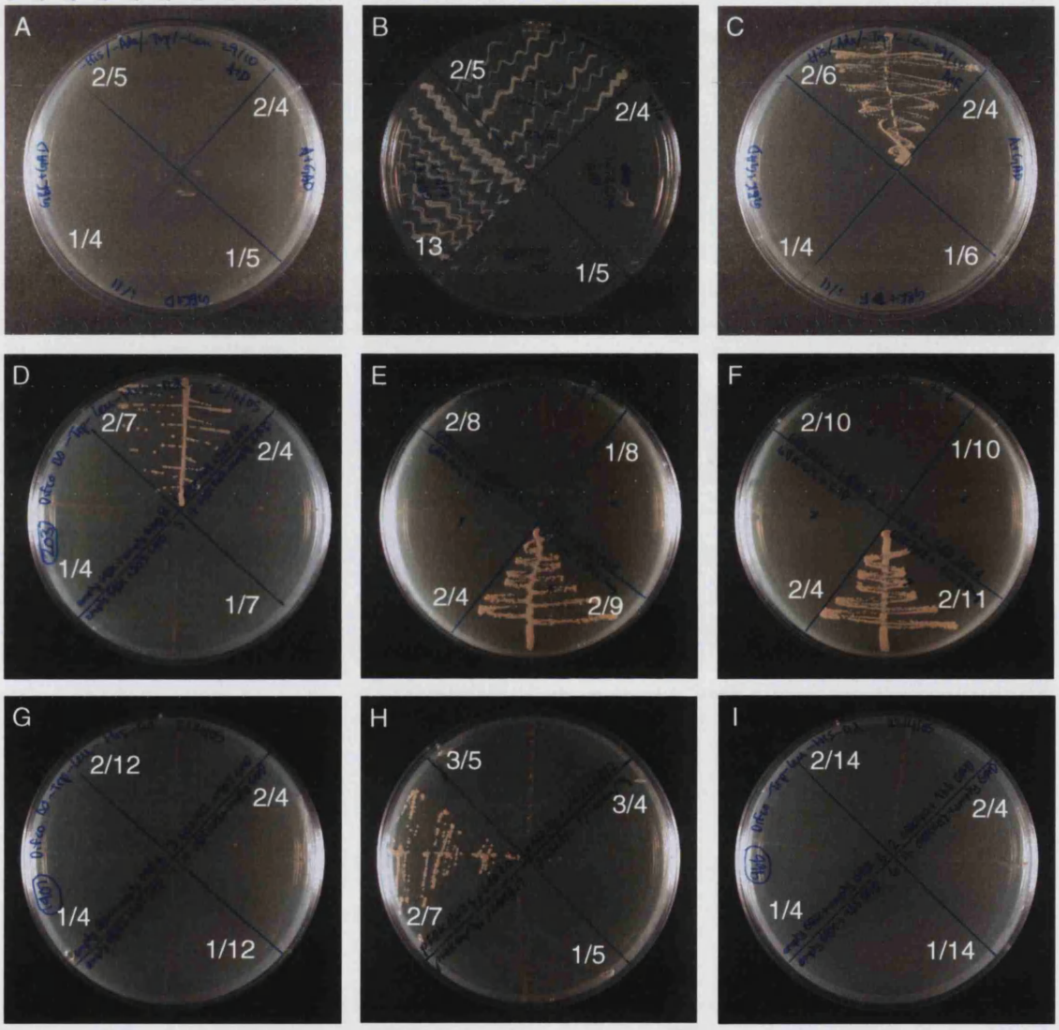


Figure 4-11 Interactions of Cdc6-1 with other full-length proteins. All plates were SD/-Trp/-Leu/-His/-Ade except B which was SD/-Trp/-Leu/-His. The plates were streaked with *S. cerevisiae* AH109 transformed with various combinations of bait and prey constructs. The bait constructs are labelled: 1) pGBKT7/empty, 2) pGBKT7/*cdc6-1* and 3) pGBKT7/*mth810*. The prey constructs are labelled: 4) pGADT7/empty, 5) pGADT7/*cdc6-1*, 6) pGADT7/*mcm*, 7) pGADT7/*mth203*, 8) pGADT7/*mth492* (full), 9) pGADT7/*mth492* (clone 42), 10) pGADT7/*mth656* (full), 11) pGADT7/*mth656* (clone 217), 12) pGADT7/*mth1907*, 13) pGBKT7/53 pGADT7/T (positive control) and 14) pGADT7/*mth966*.

Table 4-6 Summary of interactions with Cdc6-1.

Gene	Library Clone	Growth as Full-length Clone
cdc6-1	+ (48)	+*
cdc6-2	-	-
mcm	+ (67)	+
mth203	+ (11, 14)	+
mth492	+ (42)	-
mth656	+ (217)	-
mth1907	+ (26)	-
mth810	-	-
mth966	-	-

+ denotes interaction (and library clone number) and – denotes no interaction. *only grew on SD/-Trp/-Leu/-His.

Next, the extent of amino acid residues included in each of the library fragments was determined, by direct sequencing of clones, in order to gain a better insight into the protein regions involved in contact with Cdc6-1 (Table 4-7 And Fig. 4-12). Fragment 67 contained the C-terminal 243 residues of Mth1770, the MCM protein. Fragment 48 contained the C-terminal 205 residues of Mth1412, the Cdc6-1 protein. Fragments 11 and 14 contained the C-terminal 53 and 263 residues of Mth203 respectively. Fragment 42 contained the N-terminal 241 residues of Mth492. Fragment 217 contained 281 internal residues from the N-terminal portion of Mth656.

Table 4-7 Extent of residues encoded by the library fragments

	Start	End	Length
Mth203	0	426	426
11	373	426	53
14	163	426	263
Mth492	0	823	823
42	0	240	241
Mth656	0	838	838
217	144	425	281
Mth1412	0	383	383
48	178	383	205
Mth1770	0	667	667
67	423	667	243

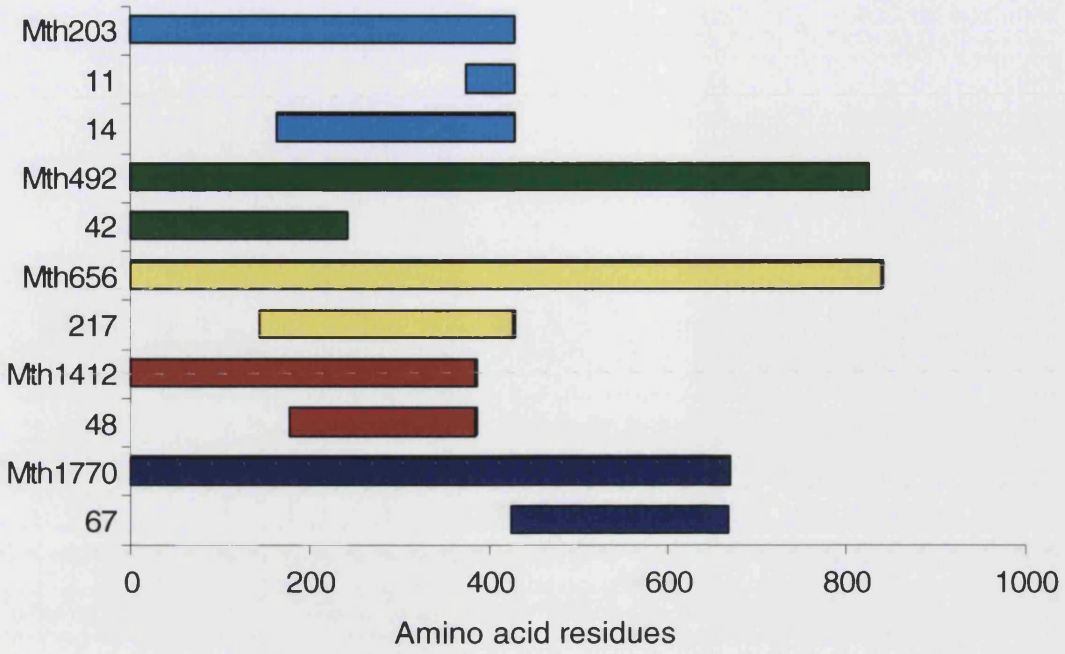
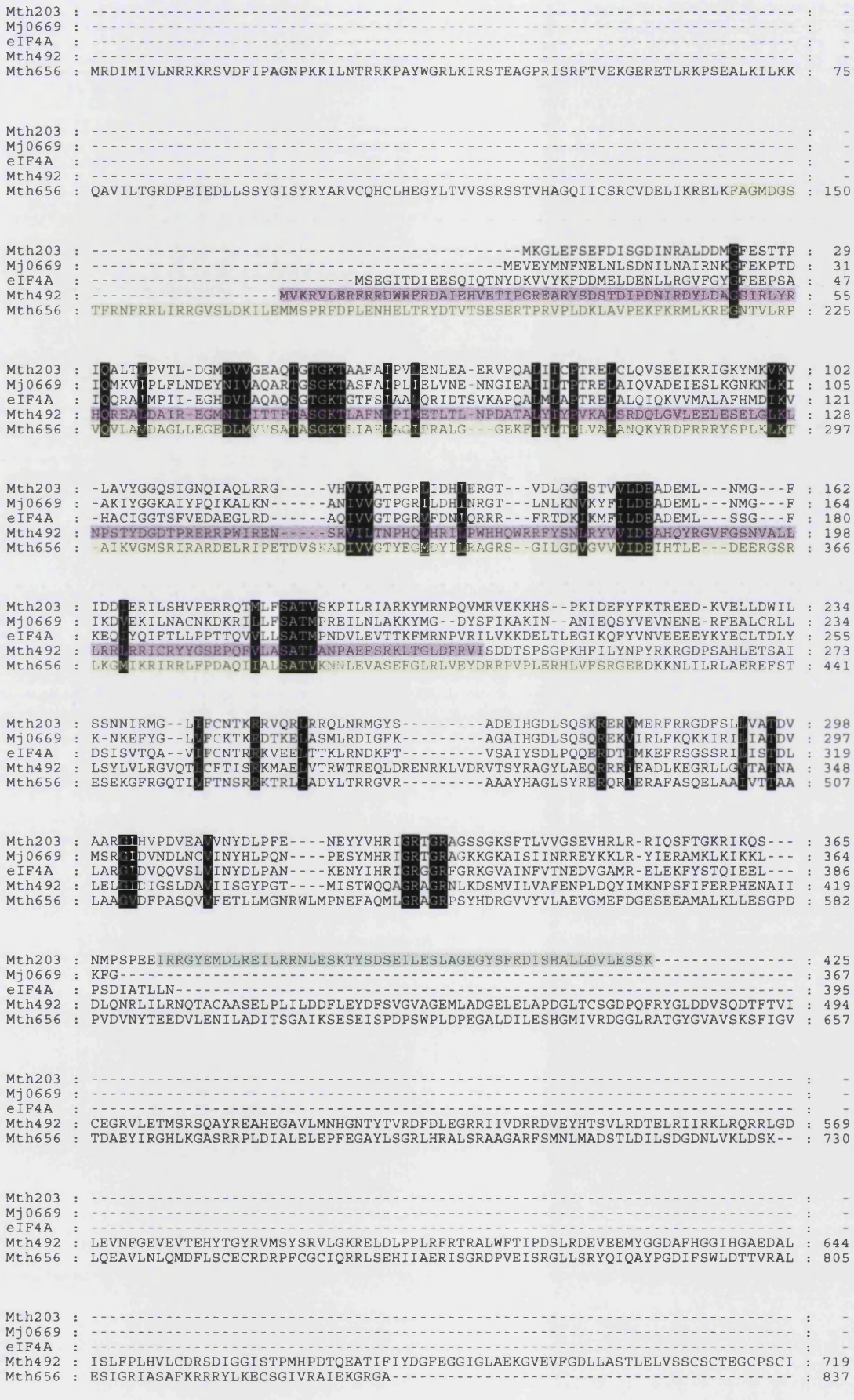


Figure 4-12 Illustration of the protein fragments that interacted with Cdc6-1. This figure shows the lengths of selected fragments identified in the Cdc6-1 interaction screen. Each fragment is aligned with its corresponding full-length protein for reference but the different proteins are not aligned with respect to each other.

The Mth203, Mth492 and Mth656 library fragments were examined more closely by highlighting the fragment residues after an alignment of the three full-length protein sequences (Fig. 4-13 next page). eIF4A and Mj0669 were included in the alignment as these proteins both represent the SF2 helicase core residues.



```
Mth203 : ------------------------------------------------------------------------ :   -
Mj0669 : ------------------------------------------------------------------------ :   -
eIF4A  : ------------------------------------------------------------------------ :   -
Mth492 : YSPKCGNENSPLHRDATVKILEYLMEQTSGEVREASEVDGDVLAGVEELCSEGNLLDAKLRLHEVLERDPGNPGA : 794
Mth656 : ------------------------------------------------------------------------ :   -


Mth203 : --------------------------- :   -
Mj0669 : --------------------------- :   -
eIF4A  : --------------------------- :   -
Mth492 : CYLMGEILRRQGEVEMAAYFHERAMNGS : 822
Mth656 : --------------------------- :   -
```

Figure 4-13 Alignment of Mth203, Mth492, Mth656, eIF4A and Mj0669 SF2 proteins. The residues contained in the Mth203, Mth492 and Mth656 library fragments are highlighted in green pink and yellow respectively. ClustalW alignment using BLOSUM series matrix and, otherwise, default parameters.

4.4 Discussion

The results presented in this chapter demonstrate that full-length MCM interacted with Cdc6-1 in a yeast two-hybrid system (Section 4.3). This observation is consistent with results from previous studies, including data from two-hybrid, affinity co-purification and far western analyses (Shin *et al.*, 2003, Kasiviswanathan *et al.*, 2005). Interaction between MCM and Cdc6 proteins has also been demonstrated for *S. solfataricus* (De Felice *et al.*, 2004b) and *P. aerophilum* (Kasiviswanathan *et al.*, 2005). However, the interaction appears to be species specific, in that MCM and Cdc6 from different species do not interact with each other (Kasiviswanathan *et al.*, 2005). Together, these findings lend support to the hypothesis that Cdc6-1 is involved in recruiting MCM to the origin of replication. MCM was also identified as a positive in the Cdc6-1 library interaction screen. This observation is consistent with the full-length interaction presented in Section 4-3.

The MCM library fragment 67, identified in the Cdc6-1 screen, corresponded to the C-terminal 243 residues (423-667) of the MCM protein. This finding is somewhat contrary to the results recently published by Kasiviswanathan and colleagues (2005). This paper outlines the cloning of specific portions of *M. thermautotrophicus* Cdc6-1 and MCM genes, followed by a directed yeast two-hybrid approach to investigate the domains required for interaction between the two proteins. They found that no binding was detected between the full-length Cdc6-1 protein and the C-terminal catalytic domain (residues 287-667) of MCM, as shown in this Chapter. However, it

is possible that the interaction between Cdc6-1 and fragment 67 (423-667) was genuine but that the interaction is abolished when the rest of the residues comprising the C-terminal domain (287-422) are included, possibly due to steric hindrance. It is also worth noting that Kasiviswanathan and colleagues (2005) reported some interaction between the winged-helix (WH) domain of Cdc6-1 and the C-terminal domain of MCM. Furthermore, full-length MCM bound to Cdc6-1 with higher affinity than the N-terminal domain alone. Secondly, it was demonstrated that domain C of the N-terminal portion of MCM (residues 92-120 and 169-286) was necessary and sufficient for interaction with Cdc6-1 (Kasiviswanathan *et al.*, 2005). That observation does not exclude the possibility that fragment 67 contains a second Cdc6-1 binding site not identified in their study. However, the question remains: why was MCM domain C not identified by the Cdc6-1 screen? Perhaps the screen was not exhaustive. If fragment 67 does reveal a novel Cdc6-1 binding site, what is its function? Perhaps binding via this site is involved in the modulation of Cdc6-1 autophosphorylation.

The Cdc6-1 WH domain corresponds to residues 287-383 and clone 48 encodes residues 178-383 (Table 4-7 And Fig. 4-12). It is therefore possible that Cdc6-1/Cdc6-1 binding is mediated by the WH domain (287-382). However, it may also require the additional upstream residues contained in fragment 48 (178-286) or they may contribute to it.

The results of the library screen suggest that thirteen different proteins interact specifically with Cdc6-1. It is unlikely that all 13 interact with Cdc6-1 *in vivo* so some are likely to be false positives. Since so little is known about how archaea control their cell cycle, none of the interactions can be immediately ruled out without further investigation. Even the less obvious candidates may have a reasonable purpose for interacting with Cdc6-1. For instance, the methyl coenzyme M reductase subunit may interact with Cdc6-1 in order to provide a means of coupling the cell cycle with methanogenesis and energy production.

Three of the genes identified in the Cdc6-1 screen (Mth203, Mth492 and Mth656) are currently annotated as "ATP-dependent RNA helicases". There is no published experimental evidence, however, that these proteins have that particular activity.

Mth203, Mth492 and Mth656 are all members of the Superfamily 2 (SF2) class of proteins. SF2 proteins have a wide range of activities and are often involved in various aspects of DNA/RNA processing (Singleton and Wigley, 2002). Certain well characterised members of SF2 do indeed exhibit RNA helicase activity *in vitro* (e.g., eIF4A) and presumably this is why Mth203, Mth492 and Mth656 have been annotated accordingly. This does not necessarily mean that Mth203, Mth492 and Mth656 have RNA specificity or even function as helicases. However, the fact that three SF2 proteins were identified by the Cdc6-1 interaction screen was significant. The initiation and elongation phases of DNA replication involve significant rearrangements of DNA structure such as: melting DNA at the origin, unwinding DNA at the replication forks, priming DNA synthesis, remodelling chromatin, etc. SF2 proteins are typically associated with processes involving rearrangements of DNA and RNA. This, therefore, made Mth203, Mth492 and Mth656 particularly interesting candidates for further study.

SF2 proteins have a conserved core structure that contains nine classical SF2 helicase motifs (Q, I, Ia, Ib, II, III, IV, V and VI). This helicase core consists of two discrete globular domains (1 and 2) that are connected by a linker region. Domain 1 contains motifs Q, I, Ia, Ib, II and III whereas domain 2 contains IV, V and VI. SF2 proteins often also have divergent N- and C-terminal extensions that determine specific characteristics such as protein-protein interactions (e.g., between the SF2 protein XPD and p44) (Coin *et al.*, 1998) and polynucleotide substrate specificities (e.g., DbpA) (Tsu *et al.*, 2001). eIF4A and Mj0669 (two model SF2 proteins) both represent the basic conserved helicase core without any extensions (Fig. 4-13).

Although fragments of Mth492 and Mth656 were identified in the Cdc6-1 screen, the full-length proteins did not interact with Cdc6-1. Mth656 has a ~175 residue N-terminal extension and a ~255 residue C-terminal extension whereas Mth492 just has a ~400 residue C-terminal extension (Fig. 4-13). The Mth492 and Mth656 library fragments are both largely within the core region of their respective proteins. Moreover, the two fragments both represent what is essentially the SF2 helicase core domain 1 of their corresponding full-length proteins. This suggests that it is domain 1 that is involved in Cdc6-1 binding. However, why was no interaction was observed in the context of full-length Mth492 and Mth656? This may be due to the

presence of the GAL4 N-terminal domain; steric hindrance may have prevented binding to Cdc6-1. Therefore, the experiments presented in this chapter do not provide enough evidence to unambiguously determine whether Cdc6-1 physically interacts with Mth492 or Mth656.

Two independent Mth203 fragments (11 and 14) were identified in the Cdc6-1 library screen. Both fragments were specific for Cdc6-1. When full-length Mth203 was tested, it too interacted specifically with Cdc6-1 (Fig. 4-11). Mth203 has a ~50 residue long C-terminal extension with respect to the SF2 helicase core (Fig. 4-13). Fragment 11 contained just the 53 C-terminal residues of Mth203, indicating that this extension was sufficient for Cdc6-1 binding. If that was the case, it would suggest that Mth203 interacts with Cdc6-1 in a completely different manner to that of Mth492 and Mth656. Furthermore, since the extension is a non-conserved portion of the protein, this particular mode of Cdc6-1 binding is likely to be specific to Mth203.

The results presented in this chapter detail the discovery of a novel interaction between Cdc6-1 and Mth203. If either protein was removed, then the reporter genes were not activated, suggesting that the interaction was specific. Furthermore, the interaction appears to be mediated by a defined region of Mth203, its 50 residue C-terminal extension. Together, these observations support the hypothesis that it is a genuine and specific interaction. However, even under these circumstances it is prudent to confirm the interaction using an independent method. Therefore, that was the aim of the next chapter.

5 Testing the Interaction between Cdc6-1 and Mth203 by Affinity Co-Purification

5.1 Introduction

The yeast two-hybrid system is known to generate false positives during library screens. The number of false positives may be kept to a minimum by using an optimised two-hybrid system, a stringent screening procedure and by careful application of controls. However, it is also prudent to verify positive two-hybrid interactions using an independent method. There are several other technologies for testing protein-protein interactions, such as: affinity co-purification (pull-down), co-immunoprecipitation, surface plasmon resonance and far-western analysis (Ausubel, 1987, Sambrook and Russell, 2001). Each of these has its own merits and drawbacks. Pull-downs are commonly used to verify two-hybrid results (Kasiviswanathan *et al.*, 2005, Marinsek *et al.*, 2006, Dionne and Bell, 2005); they are relatively simple to do and require no special equipment or custom antibodies. Briefly, a pull-down involves: producing a bait-affinity purification tag fusion protein, mixing the bait-tag and prey proteins together and then applying them to a tag-affinity column, washing the column, eluting the bait-tag and finally testing for the presence of the prey in the eluent. This assay is based on the assumption that the bait-tag and the prey stay column-bound during the wash step and that the prey elutes together with the bait-tag. Therefore, a potential limitation of this method is that it is capable of detecting only relatively stable interactions in contrast to the two-hybrid method, which is sensitive enough to detect weak and even transient interactions. Thus, some protein-protein interactions detected by yeast two-hybrid screening may not be confirmed by this method. Clearly, controls are also needed to validate the results. A negative control is used to test for non-specific binding of the prey protein to the affinity purification matrix or a non-bait protein. A positive, known interaction, control is used to check that expression, extraction, binding and washing conditions are suitable for detecting interactions.

The interaction between Cdc6-1 and Mth203 was verified by a pull-down assay. MCM was chosen as the positive control as its interaction with Cdc6-1 had been demonstrated before, both in Chapter 4 (by two-hybrid assay) and also in other

studies (by two-hybrid, far western and pull-down) (Shin *et al.*, 2003, Kasiviswanathan *et al.*, 2005). Mth810 was chosen as the negative control as it was shown not to interact with Cdc6-1 in a two-hybrid assay (Chapter 4) but, like Mth203, it is a superfamily 2 helicase. A His-tag was used to bind the Cdc6-1 bait to the affinity matrix. A S-tag was fused to each of the three prey proteins (Mth203, MCM and Mth810) to enable their detection on a Western blot. The pCDF-Duet system was used to co-express each of the three bait/prey combinations in *E. coli*. This had an added advantage over single expression in that it allowed the proteins to interact in an intracellular environment, albeit different from *M. thermautotrophicus*, before column purification.

5.2 Aims and Rationale

The aim of this chapter was to independently verify the interaction between Cdc6-1 and Mth203 as identified by yeast two-hybrid assay. In order to do that, a pull-down assay, including positive and negative controls, was used.

5.3 Results

5.3.1 Protein Expression and His-bind Column Pull-downs

The three combinations of bait and prey proteins were expressed in *E. coli* and processed through a His-bind column as described in Section 2.8. Samples from the clarified lysates and elution fractions were analysed by SDS-PAGE (Fig. 5-1). The predicted molecular weights for each of the fusion proteins are: 45.5 KDa for Cdc6-1, 50.7 KDa for Mth203, 81.1 KDa for Mth810 and 78.1 KDa for MCM. Cdc6-1 is identifiable in each gel as a strong band in lane E3 below the 50 KDa marker. Bands for the other fusion proteins (Mth203, Mth810 and MCM) are not easily distinguishable from other bands on the Coomassie stained gels.

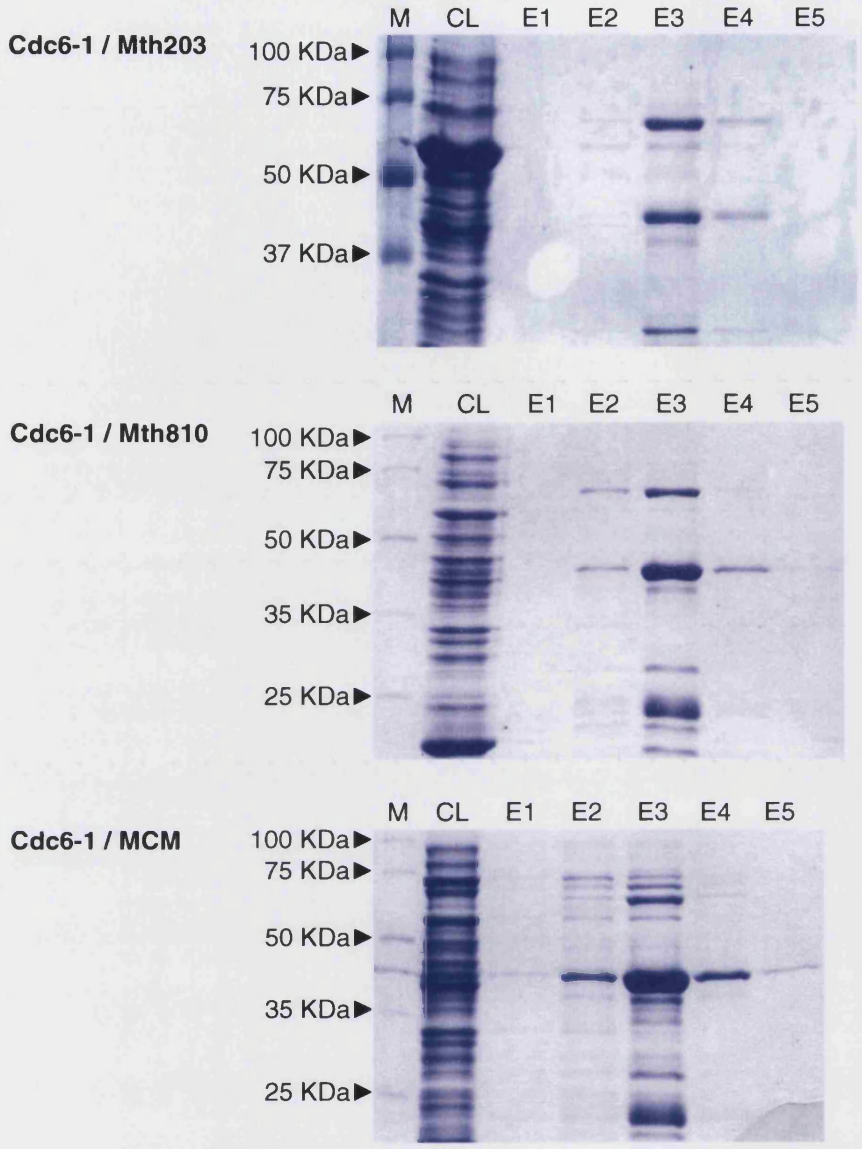


Figure 5-1 Elution profiles from the three His-bind pull-down columns. Lanes are labelled: M) markers, CL) clarified lysate and E1-E5) elution fractions 1 to 5.

5.3.2 Western Blots

An anti-His-tag Western blot confirmed that the Cdc6-1 bait was present in the clarified lysates and elution fractions of each of the three pull-downs (Fig. 5-2).

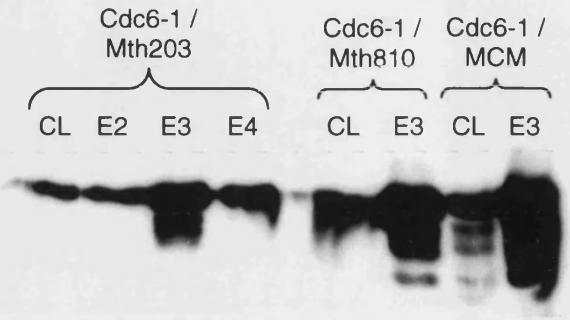


Figure 5-2 Anti His-tag western blot for the three pull-downs. CL) denotes clarified lysate and E) elution fraction. All 8 samples contained His-tagged Cdc6-1.

An S-protein Western blot confirmed that Mth203, Mth810 and MCM were present in the respective clarified lysates, before the pull-down, and therefore had been properly expressed (Fig. 5-3). The presence of Mth203 in the elution fractions (E2 and E3) of the pull-down indicated an interaction between Cdc6-1 and Mth203. Mth810 was not detected in the elution fraction suggesting that it did not interact with Cdc6-1 and was washed out of the column. As expected, MCM interacted with Cdc6-1 and, like Mth203, was present in the elution fraction.

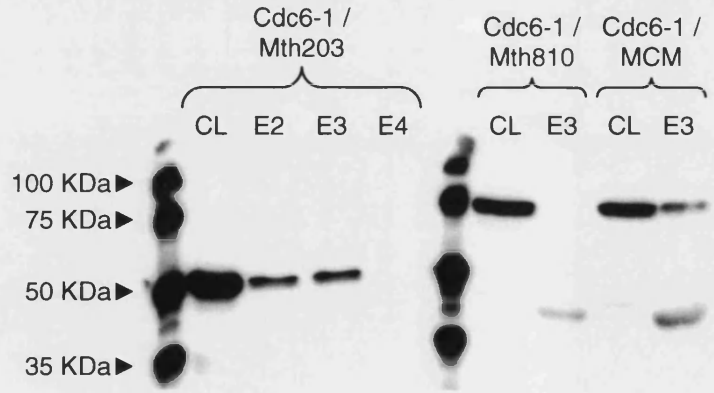


Figure 5-3 S-protein western blot for the three pull-downs. CL) denotes clarified lysate and E) elution fraction. Mth203 was present in the clarified lysate and elution fractions 2 and 3. Mth810 was present in the clarified lysate but not elution fraction 3. MCM was present in both the clarified lysate and elution fraction 3.

5.4 Discussion

The results of the pull-down assays support the case for an interaction between the *M. thermautotrophicus* proteins Cdc6-1 and Mth203. Although Mth203, Mth810 and MCM were not easily distinguished from the other bands on the Coomassie stained gels, the Westerns show that the proteins were present in the co-expression transformant cells. The premise was that if these proteins interact with Cdc6-1 then they would be co-purified with His-tagged Cdc6-1 when passed down a His-bind affinity column. It is possible that the Mth203 protein could bind non-specifically to the matrix of the His-bind affinity column and this possibility was not eliminated. Mth810 (Hel308) is a predicted superfamily 2 helicase (Guy and Bolt, 2005, Fujikane *et al.*, 2005), one of the 6 encoded by the *M. thermautotrophicus* genome, but was not identified during screening or found to interact with Cdc6-1 in a two-hybrid assay (Chapter 4). Mth810 was therefore chosen to serve as a negative control. This also had the advantage of confirming the previous yeast two-hybrid results which tested for the ability of Cdc6-1 to interact generally with superfamily 2 helicases, that is, without any biological significance. The lack of interaction between Mth810 and Cdc6-1 shown in this chapter helped to validate the pull-down approach and further supported the specific interaction between Mth203 and Cdc6-1. As a physical interaction between *M. thermautotrophicus* Cdc6-1 and MCM has already been confirmed by others (Shin *et al.*, 2003, Kasiviswanathan *et al.*, 2005), the observed interaction between Cdc6-1 and MCM in this chapter validated the pull-down method.

Lack of time prohibited repeating this experiment with the two other superfamily 2 helicases identified in the screen, Mth492 and Mth656. But, since the two full-length proteins were shown not to interact by two-hybrid assay, it is uncertain whether they would do so in a pull-down assay. However, there have been instances when proteins did not interact by two-hybrid but did interact when tested by a pull-down assay. For example, interaction was not detected between *M. thermautotrophicus* Cdc6-2 and MCM by two-hybrid (Chapter 4) (Kasiviswanathan *et al.*, 2005) but was observed by pull-down and far western analysis (Kasiviswanathan *et al.*, 2005). Therefore, given more time, performing Cdc6-1 pull-downs with Mth492 and Mth656 would be appropriate.

6 Analysis of Mth203

6.1 Introduction

Helicases are enzymes that catalyse the separation of two strands of a double helix using energy derived from the hydrolysis of nucleotides. They are found in all living organisms and participate in a wide variety of important cellular processes such as: DNA replication, recombination, repair, transcription, RNA processing and translation. Their importance to cellular function is reflected by the genomic complement of helicase genes, for example 14 are present in *E. coli* and 26 in *S. cerevisiae* (Tuteja and Tuteja, 2004).

The primary structures of helicases are characterised by having up to nine conserved motifs that cluster into a core region spanning 300 to 400 residues (Cordin *et al.*, 2006). These motifs have been conventionally named Q, I, Ia, Ib, II, III, IV, V and VI. Motifs I to VI were defined from an alignment of eight proteins involved in RNA metabolism from different species (Linder *et al.*, 1989). The Q motif was identified later from the alignment of various DEAD-box proteins (see below) (Tanner *et al.*, 2003).

Based on the occurrence and characteristics of the nine conserved motifs, helicases have been divided into five superfamilies (SFs) (Cordin *et al.*, 2006). SF1 and SF2 are the largest of these and contain at least seven of the nine motifs (I, Ia, II, III, IV, V and VI). SF3, 4 and 5 only have 2-5 conserved motifs. SF3 helicases are usually found in viruses (e.g., SV40 large T-antigen) and contain only three motifs (I, II and III). SF4 are hexameric DnaB-like helicases and contain five motifs (H1, H1a, H2, H3, and H4), of which only two (H1 and H2) are similar to conserved motifs of the other superfamilies (I and II). SF5 proteins are Rho-like helicases and also only contain motifs I and II. Other helicases, such as MCM and RuvB, have been assigned to the AAA^+ ATPase superfamily. AAA^+ proteins use energy, derived from ATP hydrolysis, to exert a mechanical force on some macromolecular substrate though they are not necessarily always associated with unwinding nucleic acids.

SF1 and SF2 helicases are structurally more closely related to each other than either is to any of the other superfamilies. They both have a conserved core fold that resembles the central region of RecA (Subramanya *et al.*, 1996). This helicase core consists of two globular domains (1 and 2) that form a deep cleft, lined with conserved residues (Singleton and Wigley, 2002). The cleft forms the nucleotide binding site whereas the DNA/RNA substrate interacts with the base of the helicase core. SF1 and SF2 helicases are active as monomers or dimers in contrast with the other superfamilies, which mostly form active hexameric complexes (Cordin *et al.*, 2006). SF1 helicases contact the bases of the nucleic acid substrate and translocate along single stranded DNA/RNA whereas SF2 helicases contact the phosphodiester backbone and may therefore translocate along single or double stranded DNA/RNA (Singleton and Wigley, 2002). While the structure of the helicase core is well conserved, many SF1 and SF2 proteins also contain variable N- and C-terminal extensions (although some SF1 helicases may have variable inserts in the core e.g., PcrA) (Singleton and Wigley, 2002). These extra variable domains are needed to tailor the helicase core to a specific function by, for example, targeting it to a particular substrate or protein complex.

Not all proteins that possess SF1 or SF2 helicase motifs actually have strand separation functionality. For example, UvrB (a SF2 protein) translocates along dsDNA, without unwinding it, until it encounters a lesion, then it recruits UvrC to initiate DNA repair at that site. Therefore, it has been suggested that the helicase motifs don't necessarily imply unwinding *per se*, merely active translocation along nucleic acids (Singleton and Wigley, 2002).

DEAD-box proteins form the largest family within the SF2 helicases. The motifs that are characteristic of the DEAD-box family have been found in over 500 proteins (Cordin *et al.*, 2006). DEAD-box proteins are distinguished from other SF2 families (DEAH, DExH, etc.) on the basis of differences in their nine conserved motifs. For example, the presence of the entire Q motif is highly characteristic of the DEAD-box family but not the other SF2 helicases. The name of the family was derived from the residue sequence (DEAD) of its motif II.

Motifs I and II, also known as Walker A and B, bind the phosphates of ATP through a coordinated Mg^{2+} ion (Rocak and Linder, 2004). All helicases have the Walker A and B motifs, which are also found in many other ATPases in general. Motif III couples ATP hydrolysis with the conformational changes involved in nucleic acid unwinding / translocation (Rocak and Linder, 2004). Motifs Q and VI are also involved in ATP binding and hydrolysis, whereas Ia, Ib, IV, and V are probably involved in polynucleotide substrate binding (Rocak and Linder, 2004).

eIF4A is the prototypical DEAD-box protein. It represents the helicase core structure as it has no N- and C-terminal variable regions. Crystal structures have been solved for domains 1 and 2 separately and then combined (Caruthers *et al.*, 2000). eIF4A functions as part of a multi-protein complex with eIF4B and eIF4H, both of which are necessary for its activity. This complex unwinds the 5' untranslated region of eukaryotic mRNA to prepare it for translation. The crystal structure has also been solved for a *M. jannaschii* eIF4A homologue (Fig. 6-0) (Story *et al.*, 2001). Mj0669, unlike eIF4A, forms a dimer but its function / helicase activity has not yet been characterised.

Figure 6-0 The Mj0669 monomer protein structure. Mj0669 has a dumbbell shaped structure with globular N- and C-terminal domains connected by a linker shown in the middle of the figure. The N- and C-termini are marked on the figure (Story *et al.*, 2001).

Most of the characterized DEAD-box proteins are RNA helicases. However, examples of DEAD-box DNA helicases have been identified. For example, *S. cerevisiae* Dbp9p (Kikuma *et al.*, 2004) and *Plasmodium falciparium* PFDH60 (Pradhan *et al.*, 2005) have both been shown to unwind dsDNA. This suggests that motif characteristics, alone, are not sufficient to predict a specific preference for DNA or RNA substrates. DEAD-box proteins generally have no base sequence specificity owing to the fact that the helicase core contacts the phosphodiester backbone. A notable exception is DbpA, which has a C-terminal variable domain that binds to a specific 23S rRNA sequence (Tsu *et al.*, 2001).

Relatively few predicted DEAD-box proteins manifest helicase activity *in vitro*. In cases where helicase activity has been tested for, but not detected, this may be due to a requirement for: post-translational modification, a specific substrate (e.g., as with DbpA) or additional protein cofactors (e.g., as with eIF4A). However, most DEAD-box proteins do exhibit ATPase activity *in vitro* (though this is generally low compared to other ATPases) and this activity is often stimulated by RNA. Preferences for particular NTPs differ between DEAD-box and DEAH-box families: DEAD-box proteins tend to require ATP, whereas DEAH-box proteins tend to be more promiscuous in their use of NTPs (Rocak and Linder, 2004). Reported DEAD-box Michaelis constants (K_m) range from 80 μM to 1 mM ATP (Cordin *et al.*, 2006). While this shows a relatively low affinity for ATP (in comparison, SecA is in the nM range) it is still well below intracellular ATP concentrations (typically 5 to 10 mM) (Rocak and Linder, 2004). DEAD-box turnover numbers (k_{cat}) are also low, ranging from 3 min^{-1} (mammalian eIF4A) to 600 min^{-1} (DbpA) (Lorsch and Herschlag, 1998, Tsu *et al.*, 2001) . Therefore, caution is required when drawing conclusions from *in vitro* studies without analyses including relevant protein cofactors. It is clear that DEAD-box proteins often act in concert with other proteins to carry out their specific functions. Furthermore, inclusion of these accessory proteins into *in vitro* studies may yield K_m and K_{cat} values that better resemble the true *in vivo* activities.

6.2 Aims and Rationale

The aim of this chapter was to begin an investigation into the function of Mth203. The first objective was to study the primary sequence. The second objective was to develop methods to express and purify Mth203 protein. The third objective was to develop and use an assay to detect and measure Mth203's ATPase activity. The last objective was to investigate the affect of various conditions on this activity.

6.3 Results

6.3.1 Mth203 Sequence Analysis

The Mth203 primary sequence exhibits all nine conserved DEAD-box SF2 helicase motifs and their characteristic spacing (Fig. 6-1).

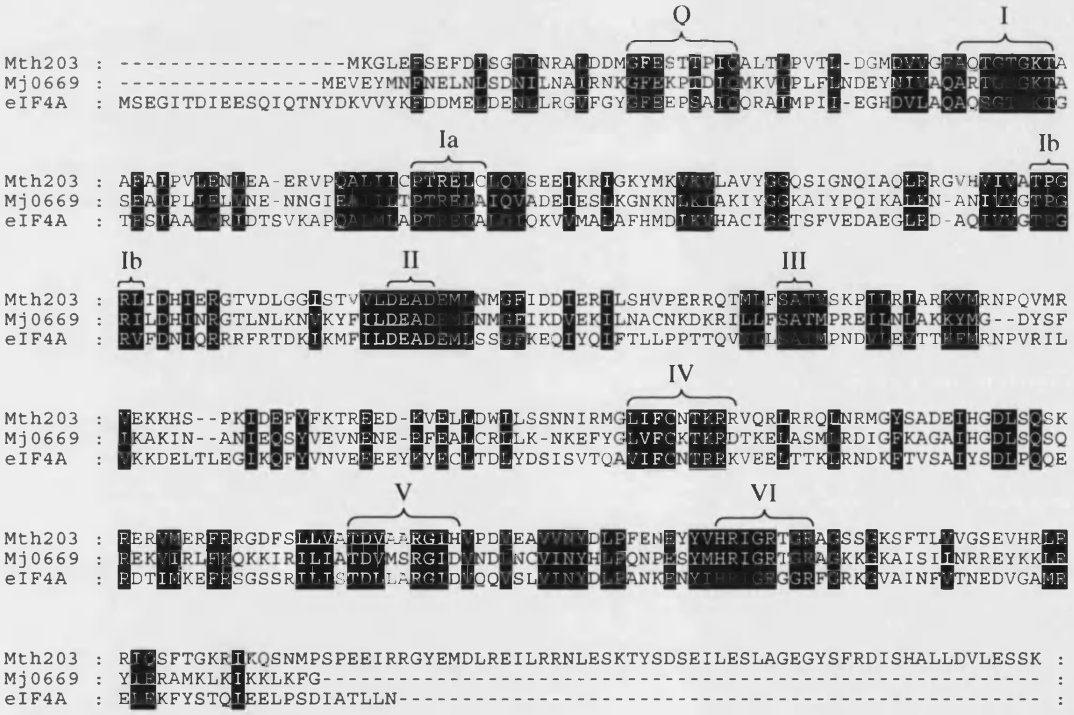


Figure 6-1: Alignment of Mth203, Mj0669 and eIF4A DEAD-box proteins. ClustalW alignment using BLOSUM series matrix and, otherwise, default parameters. Positions of the nine conserved motifs are labelled on the alignment. Labelling is according to the convention of Cordin *et al.*, 2006.

A closer look at the Mth203 conserved motifs (Table 6-1) shows that they conform well to the consensus sequences. However, there are a few notable substitutions. The Q motif proline is substituted with a threonine. The proline residue may normally be involved in an interaction with the conserved upstream phenylalanine (Cordin *et al.*, 2006). Also, the Ia motif alanine is substituted with cysteine and, lastly, the last aspartate of motif V is replaced with a histidine residue. However, it is not obvious whether these amino acid differences have a significant affect on Mth203's overall structure or function.

Table 6-1: Comparison between the conserved motifs of Mth203, Mj0669 and eIF4A.

Motif		Q	I	Ia	Ib
Consensus	F	GaccPoh1Q	AxTGoGKT	PTRELA	TPGR1
Mth203	F	GFESTTPIQ	AQTGTGKT	PTRELC	TPGRL
Mj0669	F	GFEKPTDIQ	ARTGSGKT	PTRELA	TPGRI
eIF4A	F	GFEEPSAIQ	AQSGTGKT	PTRELA	TPGRV

Motif	II	III	IV	V	VI
Consensus	DEAD	SAT	1IFhxT+cx	TDVuARGID	HRIGRTGR
Mth203	DEAD	SAT	LIFCNTKRR	TDVAARGIH	HRIGRTGR
Mj0669	DEAD	SAT	LVFCKTKRD	TDVMSRGID	HRIGRTGR
eIF4A	DEAD	SAT	VIFCNTRRK	TDLLARGID	HRIGRGGR

Standard single letter code and additional symbols: x = any; a = F, W, Y; c = D, E, H, K, R; o = S, T; h = A, F, G, I, L, M, P, V, W, Y; 1 = I, L, V; + = H, K, R; u = A, G. Residues that do not conform to the consensus sequence are underlined. Consensus sequence from (Cordin *et al.*, 2006).

Mth203 has a ~50 residue long C-terminal extension (Fig. 6-1). N- and C-terminal variable regions are often present in DEAD-box proteins and are significant in augmenting their function. It is possible that the Mth203 C-terminal extension is involved in the interaction with Cdc6-1 or a specific nucleic acid substrate or in dimerisation. The fact that this region is present in the truncated library fragment (Fig. 4-13) supports the possibility of its involvement in the interaction with Cdc6-1.

This hypothesis could be assessed by testing a Mth203 C-terminal truncation mutant in a yeast two-hybrid assay with Cdc6-1.

The top five closest proteins (assessed by GenTHREADER) to Mth203's predicted structure are all SF2 helicases (Table 6-2). However, they consist of: three RNA helicases (DjVLGB, BstDEAD and eIF4A), a DNA helicase (RecQ) and a protein involved in protein translocation (SecA).

Table 6-2: Mth203 GenThreader results.

Name	PDB id.	P-value	Title
DjVLGB	1wrb	4×10^{-11}	Crystal structure of the N-terminal RecA-like domain of a DEAD-box RNA helicase, the *Dugesia japonica* vasa-like gene B protein (Kurimoto *et al.*, 2005)
BstDEAD	1q0u	5×10^{-11}	Crystal structure of the BstDEAD N-terminal domain: a novel DEAD protein from *Bacillus stearothermophilus* (Carmel and Matthews, 2004)
eIF4A	1qva	5×10^{-11}	Crystallographic structure of the amino terminal domain of yeast initiation factor 4A, a representative DEAD-box RNA helicase (Johnson and McKay, 1999)
SecA	1m6n	6×10^{-11}	Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA (Hunt *et al.*, 2002)
RecQ	1oyw	1×10^{-10}	High-resolution structure of the *E. coli* RecQ helicase catalytic core (Bernstein *et al.*, 2003)

Top five similar protein structures from a Mth203 GenThreader search using default parameters (McGuffin *et al.*, 2000).

The top five hits from a BLASTP search using Mth203 primary sequence as a query were all superfamily 2 DNA/RNA helicases (Table 6-3). They were all predicted proteins with no published experimental evidence. Whereas Mth203 is 425 residues long, the BLASTP hits range from 521 to 575 residues.

Table 6-3: Mth203 BLASTP results.

Locus tag	organism	annotation	e-value
Msp_1228	*Methanosphaera stadmanea* DSM3091	Predicted helicase	2×10^{-106}
CphyDRAFT_3519	*Clostridium photofermentans* ISDg	DbpA RNA-binding; DEAD/DEAH-box helicase like	6×10^{-102}
Dde_0152	*Desulfovibrio desulfiricans* G20	DEAD/DEAH-box helicase like	2×10^{-101}
ABC_2200	*Bacillus clausii* KSM-K16	ATP-dependent RNA helicase	2×10^{-100}
LA4049	*Leptospira interrogans* serovar Lai strain 56601	ATP-dependent RNA helicase	1×10^{-99}

Top five similar protein structures from a Mth203 BLASTP search of the GenBank non-redundant CDS database using default parameters.

Therefore, from the perspective of primary and tertiary structure similarity searches, it is clear that Mth203 belongs to the SF2 class of proteins although its substrate specificity cannot be inferred.

Mth203 is not part of an operon and the genes in the vicinity do not appear to have related functions (see Table 6-4). However, it may be worth noting that Mth208 encodes a subunit of the *M. thermautotrophicus* family B polymerase. Mth203's function cannot be inferred from its genomic context.

Table 6-4: Genomic context of Mth203.

Mth ORF Code	Mth Genome Annotation*
200	Cobalamin biosynthesis protein M related protein
201	Hypothetical protein
202	Hypothetical protein
203	ATP-dependent RNA helicase, eIF-4A family
204	Hypothetical protein
205	Hydrogenase expression / formation protein HypE
206	Hypothetical protein
207	Ribosomal protein S8
208	DNA-dependent DNA polymerase family B (PolB2)

*Annotation taken from *M. thermautotrophicus* complete genome gi15678031 (Smith *et al.*, 1997)

6.3.2 Mth203 Expression and Purification

Mth203 was expressed in *E. coli* to produce a stock of protein for an *in vitro* analysis. The His-tagged Mth203 fusion protein had a predicted molecular mass of 50.7 KDa. It was expressed in a 400 ml culture as described in Section 2.9. Expression was monitored by taking samples and analysing them by SDS-PAGE (Fig. 6-2).

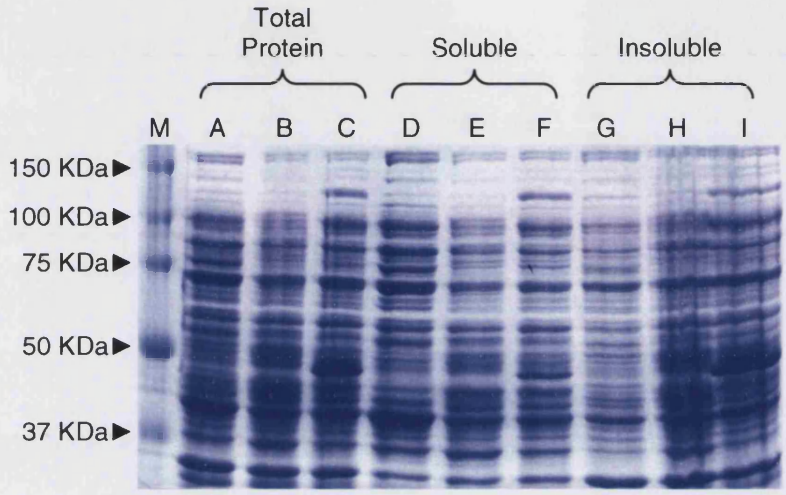


Figure 6-2 Expression of Mth203 in *E. coli*. Mth203 protein was expressed in an *E. coli* culture (Section 2.9). An aliquot of culture was taken during mid-log phase, before induction (lanes A, D and G). A second aliquot was taken, after overnight incubation, from an un-induced control culture (lanes B, E and H). Finally, a third aliquot was taken from the IPTG-induced culture (lanes C, F and I). Each sample was processed to give further samples containing: total cellular protein (TCP) (lanes A, B and C), soluble (lanes D, E and F) and insoluble (lanes G, H and I) material. Mth203 (predicted mass 50.7 KDa) was successfully over-expressed in the induced culture (C). Much of it was insoluble (I) but some remained soluble (F).

The Mth203 expression culture was harvested and processed, as described in Section 2.9, to give ~15 ml of clarified lysate (lane F in Fig. 6-2). Mth203 was then purified through a His-bind column and eluted with imidazole in a series of eight 0.5 ml fractions (Fig. 6-3). The third and fourth fractions, which contained most of the Mth203 protein, were pooled to give one sample for the next stage.

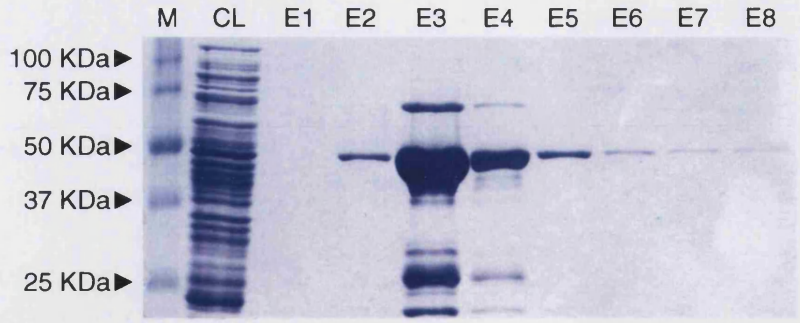


Figure 6-3 Elution profile of Mth203 from His-bind column. The lanes are labelled: M) markers, CL) clarified lysate and E1-E8) elution fractions 1 to 8. Most of the Mth203 protein eluted in fractions 3 and 4.

Mth203 was further purified by size exclusion chromatography (Section 2.9). To obtain an estimate for the molecular mass of Mth203 in its native conformation, the size exclusion column was calibrated with molecular mass standards (Fig. 6-4).

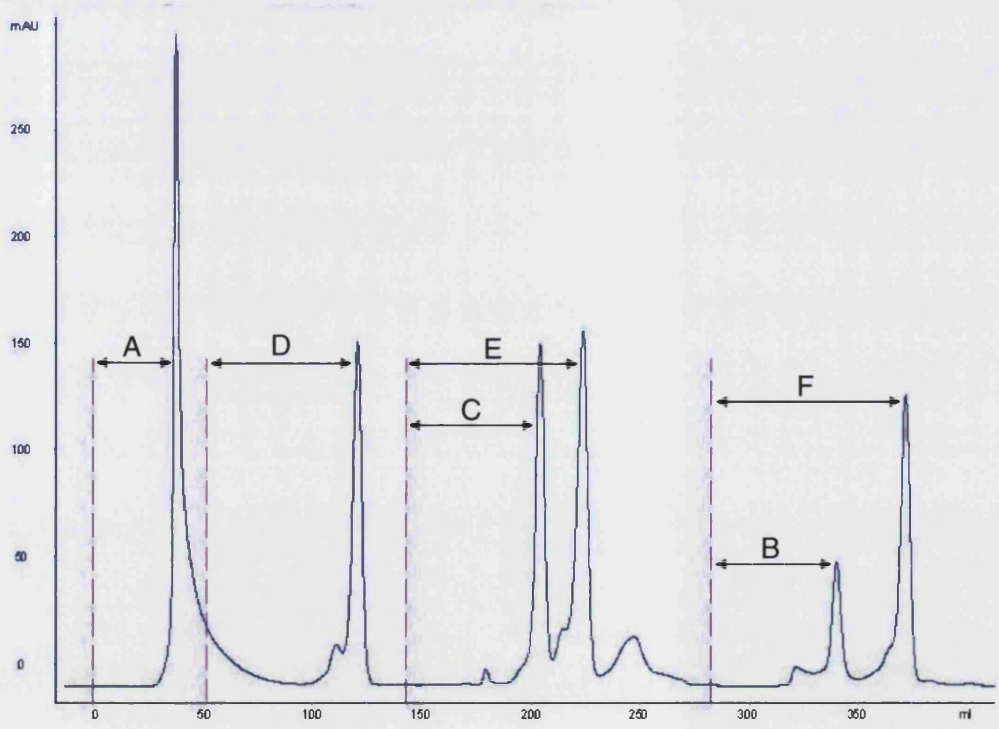


Figure 6-4 A_{280} trace of effluent from size exclusion column showing molecular mass standards. The interval between point of application (dashed line) and elution for each standard is illustrated by a double headed arrow. In order of molecular mass (largest to smallest): A) blue dextran, B) β-amylase, C) alcohol dehydrogenase, D) BSA, E) carbonic anhydrase and F) cytochrome c.

First, the blue dextran standard was used to determine the void volume (V_o) of the column (37.75 ml). Next, the elution volume (V_e) was determined for each of the protein standards by accurately measuring the volume of effluent collected between the point of each sample application and the maximum point of its corresponding elution peak. The relative elution volume (V_e/V_o) was then calculated for each standard (Table 6-5).

Table 6-5 Information from molecular mass standards.

Name	elution volume V_e (ml)	V_e/V_o	molecular mass (KDa)
A) Blue Dextran	37.75	1	2000
B) β-Amylase	57.69	1.528	200
C) Alcohol Dehydrogenase	61.9	1.64	150
D) BSA	69.24	1.834	66
E) Carbonic Anhydrase	81.65	2.163	29
F) Cytochrome C	89.36	2.367	12.4

Log molecular mass versus V_e/V_o was then plotted giving a standard curve (Fig. 6-5).

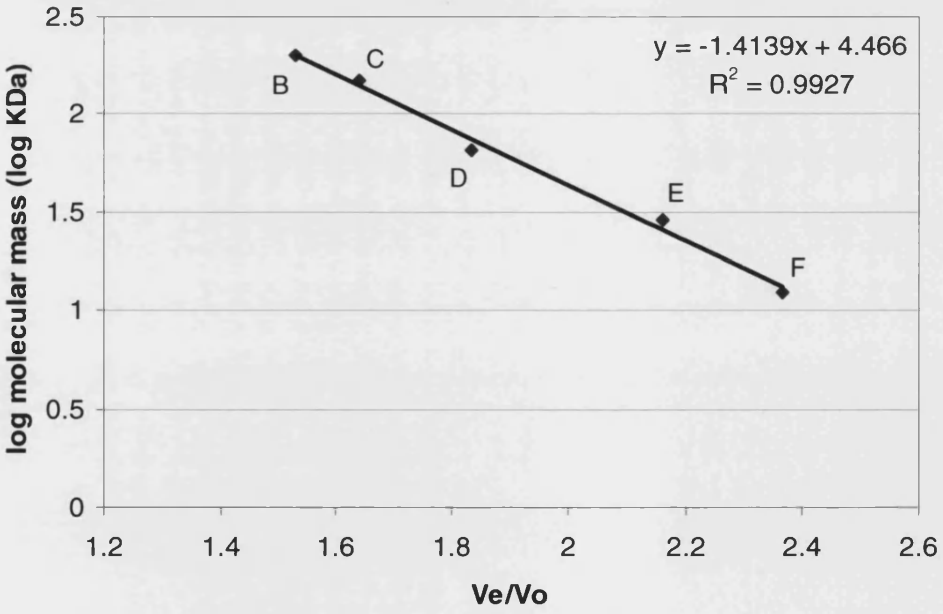


Figure 6-5 Calibration of size exclusion column using molecular mass standards. Five standards (B to F in Table 6-5) were plotted on a graph of $\log_{10}$ molecular mass against V_e/V_o. The R^2 correlation coefficient value and equation of the regression line are displayed in the top right-hand corner.

When the His-bind column purified Mth203 sample was applied to the size exclusion column, the peak corresponding to Mth203 had a V_e/V_o of 64.34 / 37.75 = 1.704 (Fig. 6-6).

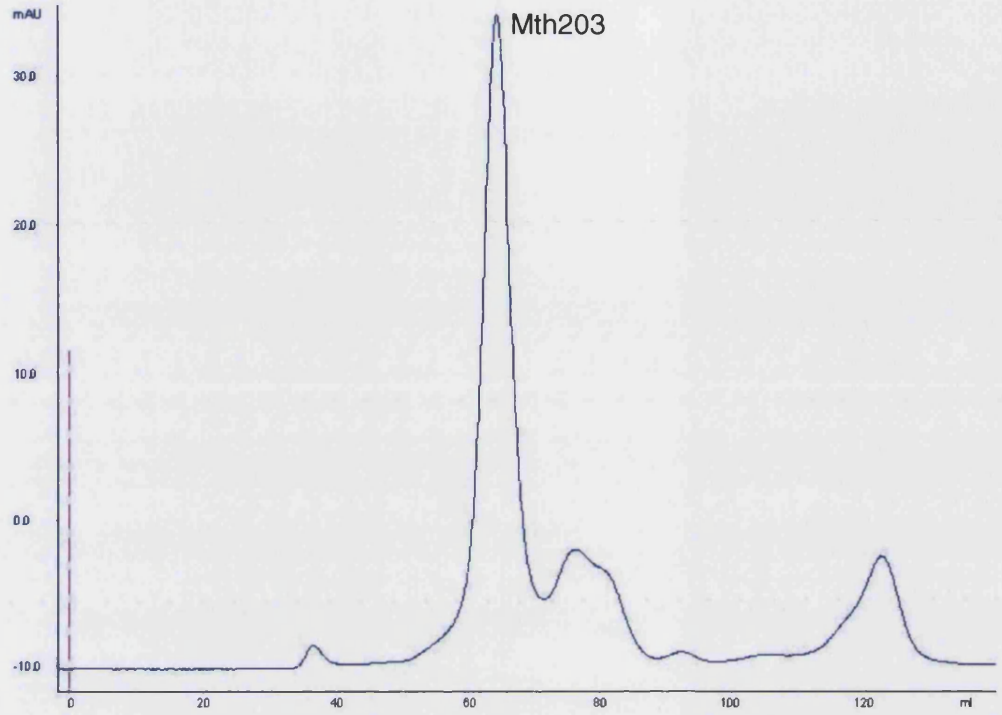


Figure 6-6 A_{280} trace showing the elution profile of Mth203 from size exclusion column. The peak that corresponded to Mth203 had an elution volume of 64.34 ml.

This corresponded to a molecular mass of 113.8 KDa. Given that the predicted molecular mass of Mth203 is 50.7 KDa, this observation suggests that the native protein associates as a dimer.

Interestingly, the Mj0669 protein also exists as a dimer, in crystal, although no other helicases have been observed to crystallise as dimers, including the closely related eIF4A (Story *et al.*, 2001). Story and co-workers (2001) also identified a motif (YSF) located 16 residues downstream from the Mj0669 motif III that contributes aromatic residues to the dimer interface. However, no such similar (aromatic-x-aromatic) motif is present in the equivalent region of the Mth203 sequence. Therefore, dimerisation of Mth203 may involve non-homologous region(s) of the protein.

Fractions from the Mth203 size exclusion peak were pooled, concentrated and stored at -20 °C (Section 2.9). A sample of this Mth203 stock was then analysed by SDS-PAGE (Section 2.3.13) to asses its purity and estimate the molecular mass of denatured Mth203 (see Fig. 6-7).

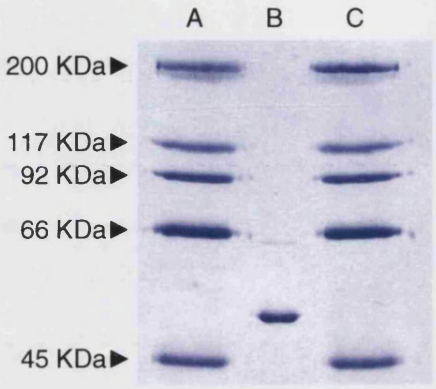


Figure 6-7 Purified Mth203 protein stock. 10% SDS-PAGE gel loaded with Mth203 (lane B) flanked by protein molecular weight markers (lanes A and C).

The distance migrated by each band was measured (from the bottom of the well) for each of the proteins in Fig. 6-7 (see Table 6-6) and a log molecular mass versus migration distance standard curve plotted (Fig. 6-8).

Table 6-6 Information on SDS-PAGE protein molecular mass markers

	distance migrated (mm)				
Name	lane A	lane B	lane C	mean	molecular mass (Kda)
A) 200	9	-	9	9	200
B) 117	19	-	19.5	19.25	117
C) 92	23	-	23.5	23.25	92
D) 66	30	-	30.5	30.25	66
E) 45	47	-	47	47	45
Mth203	-	41	-	41	-

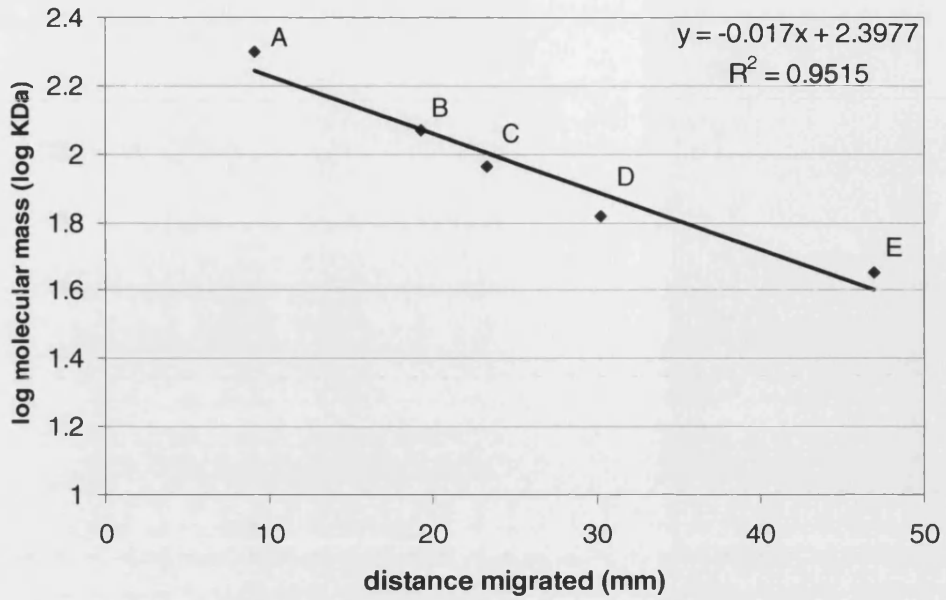


Figure 6-8 Calibration of SDS-PAGE standards for estimating the molecular mass of Mth203. Five protein standards (A to E in Table 6-6) were plotted on a graph of $\log_{10}$ molecular mass against distance migrated. The R^2 correlation coefficient value and equation of the regression line are displayed in the top right-hand corner.

Using this standard curve, the Mth203 band corresponded to a molecular mass of 50.2 KDa. This observation corroborates the predicted molecular mass of Mth203 (50.7 KDa) based on its primary sequence and also indicates that the correct size exclusion fractions were used. The protein concentration of a sample of the Mth203 stock was measured by Bradford assay (Section 2.3.14) to be 180 μg/ml, which is equivalent to a Mth203 monomer concentration of 3.6 μM.

6.3.3 Mth203 ATPase Assays

The malachite green assay was firstly calibrated using samples that contained known quantities of potassium phosphate (Fig. 6-9). The equation for the regression line for this standard curve was used to convert A_{630} readings into quantities of inorganic phosphate (Pi) in nmols.

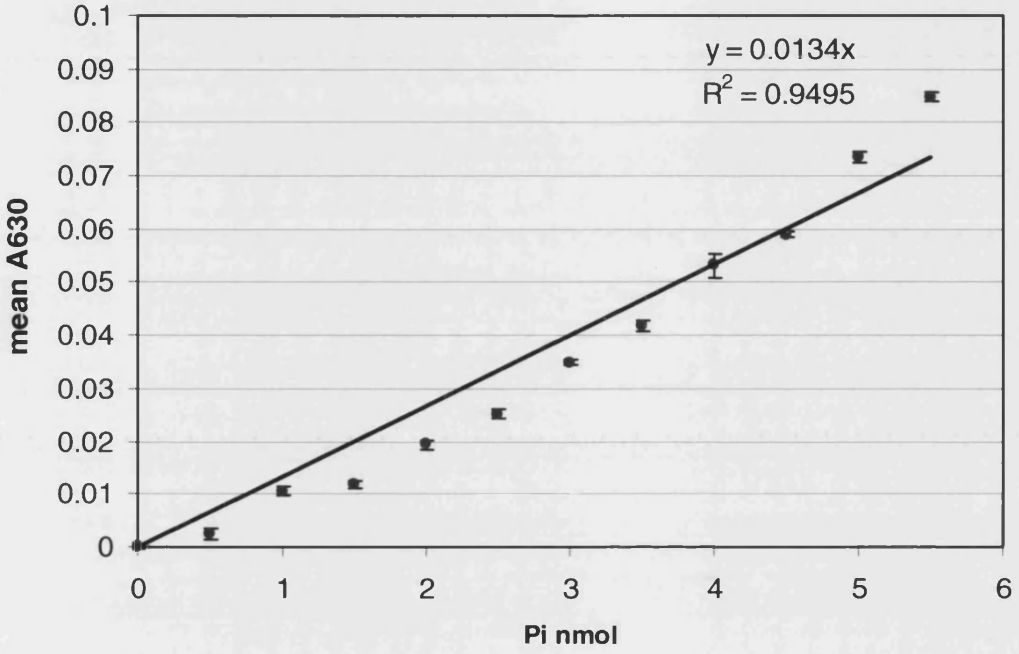


Figure 6-9 Calibration of malachite green assay with a range of phosphate concentrations. Solutions of potassium phosphate were prepared to give a range of concentrations. Five replicate 50 µl samples of each concentration were assayed with malachite green (Section 2.11). Mean A_{630} was plotted against the quantity of Pi in a 50 µl sample. A regression line was then calculated (intercept set to 0).

Throughout, standard reaction conditions (Section 2-10) were used (unless otherwise stated) and 50 µl samples were taken at each time point for malachite green analysis. Every ATPase assay was performed alongside the relevant no-enzyme control, to which enzyme storage buffer (Section 2.9) was added in place of Mth203 enzyme stock. A reading was taken for every reaction at 0 hr and then subtracted from all subsequent readings for that reaction in order to control for background. Boiling Mth203 stock abolished ATPase activity (Fig. 6-10) confirming that the activity was heat denaturable and most likely due to the Mth203 protein.

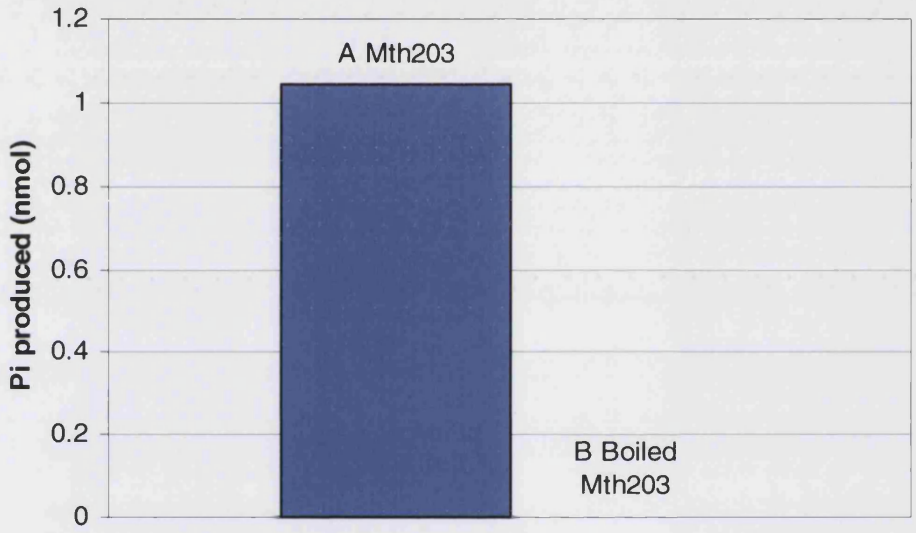


Figure 6-10 Effect of boiling Mth203 on its ATPase activity. Boiling a sample of Mth203 for 3 min abolished any observable ATPase activity in a 2 hr malachite green assay.

The amount of Pi released was measured at different time-points for 4 different concentrations of Mth203 (Fig. 6-11A). Each increase in Mth203 concentration gave a corresponding increase in Pi production. The graphs curve slightly upward suggesting a slight increase in the rate of reaction with time. The average rate of reaction was plotted for each Mth203 concentration (Fig. 6-11B). With only four data points it is hard to tell, but this appears to give a straight line. This suggests that the substrate is in excess and confirms that the observed ATPase activity was a consequence of protein input into the assay.

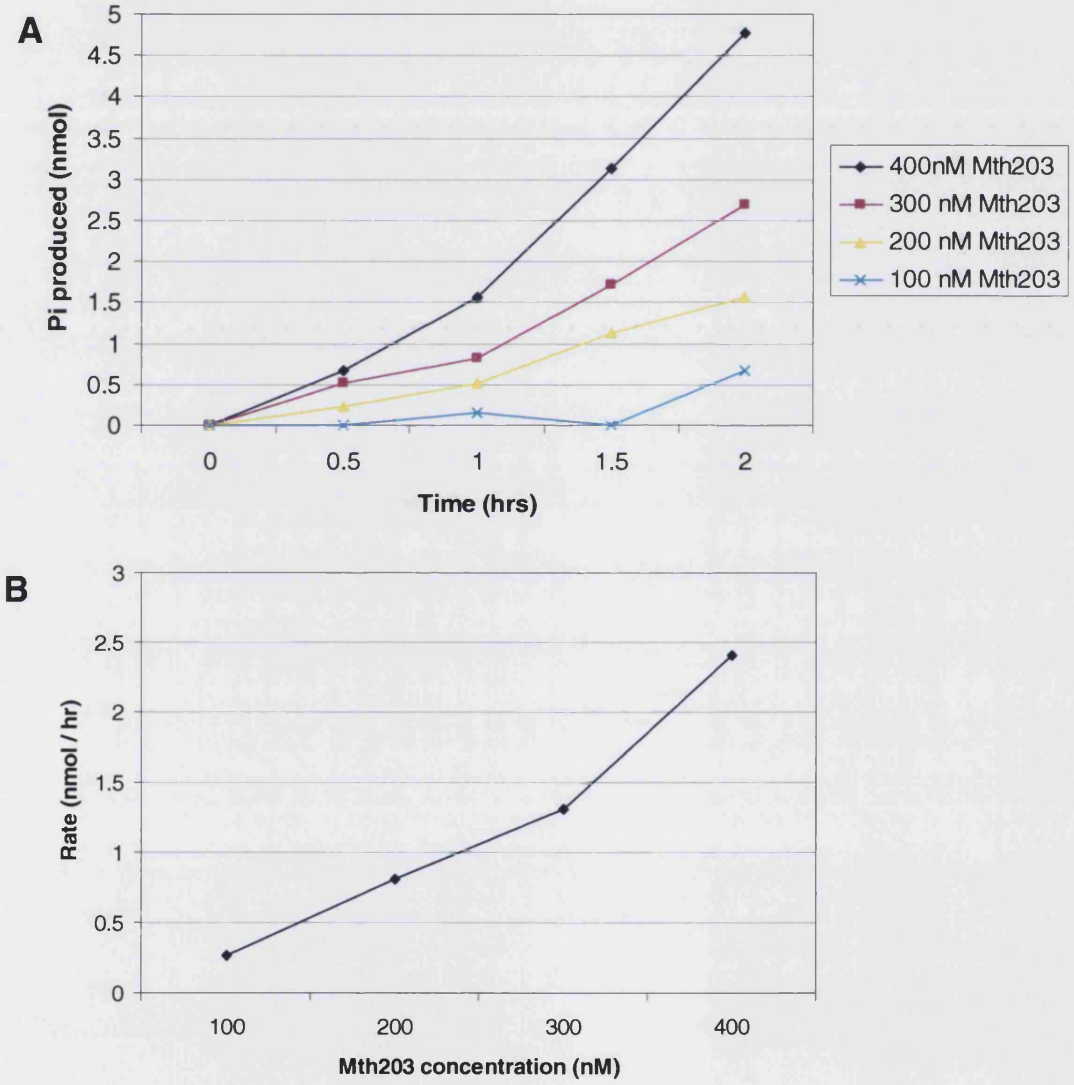


Figure 6-11 Effect of varying Mth203 concentration. Four reactions were set up containing different concentrations of Mth203. 50 µl aliquots were taken every half hour and assayed for Pi (A). The rate of each reaction was estimated by calculating the regression line and was plotted against Mth203 concentration (B).

Replacing Mg^{2+} with Mn^{2+} at varying concentrations did not significantly alter the ATPase activity of Mth203 under the conditions tested (Fig. 6-12). Clearly, though, a source of divalent cation was necessary for ATPase activity as demonstrated by the milliQ and EDTA negative controls. This is supporting evidence that the activity was due to Mth203 as DEAD-box proteins are known to require an active site Mg^{2+} cofactor. However, Mg^{2+} could be replaced with Mn^{2+}.

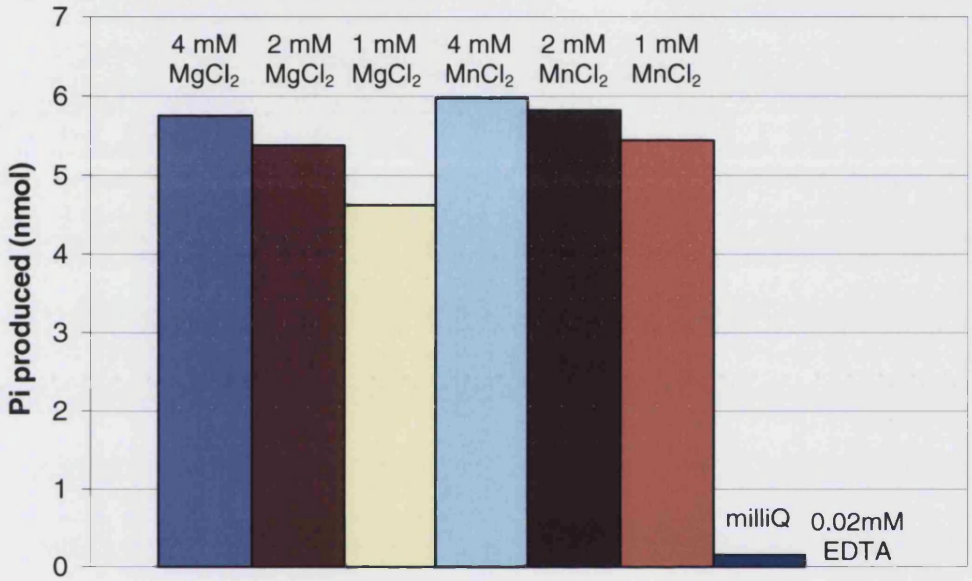


Figure 6-12 Effect of changing the divalent cation. Mth203 reactions were set up containing different concentrations of $MgCl_2$ or $MnCl_2$ and assayed for Pi after 4 hrs. ATPase activity was slightly higher in reactions containing $MnCl_2$ than $MgCl_2$.

Mth203 ATPase activity was tested at a range of temperatures and 50°C was identified as the optimum (Fig. 6-13). The optimum temperature for *M. thermautotrophicus* growth is 65 °C. A more precise estimate of Mth203's optimum could be obtained by increasing the granularity of temperatures tested. Also, the apparent optimum temperature would increase if a shorter incubation period was used.

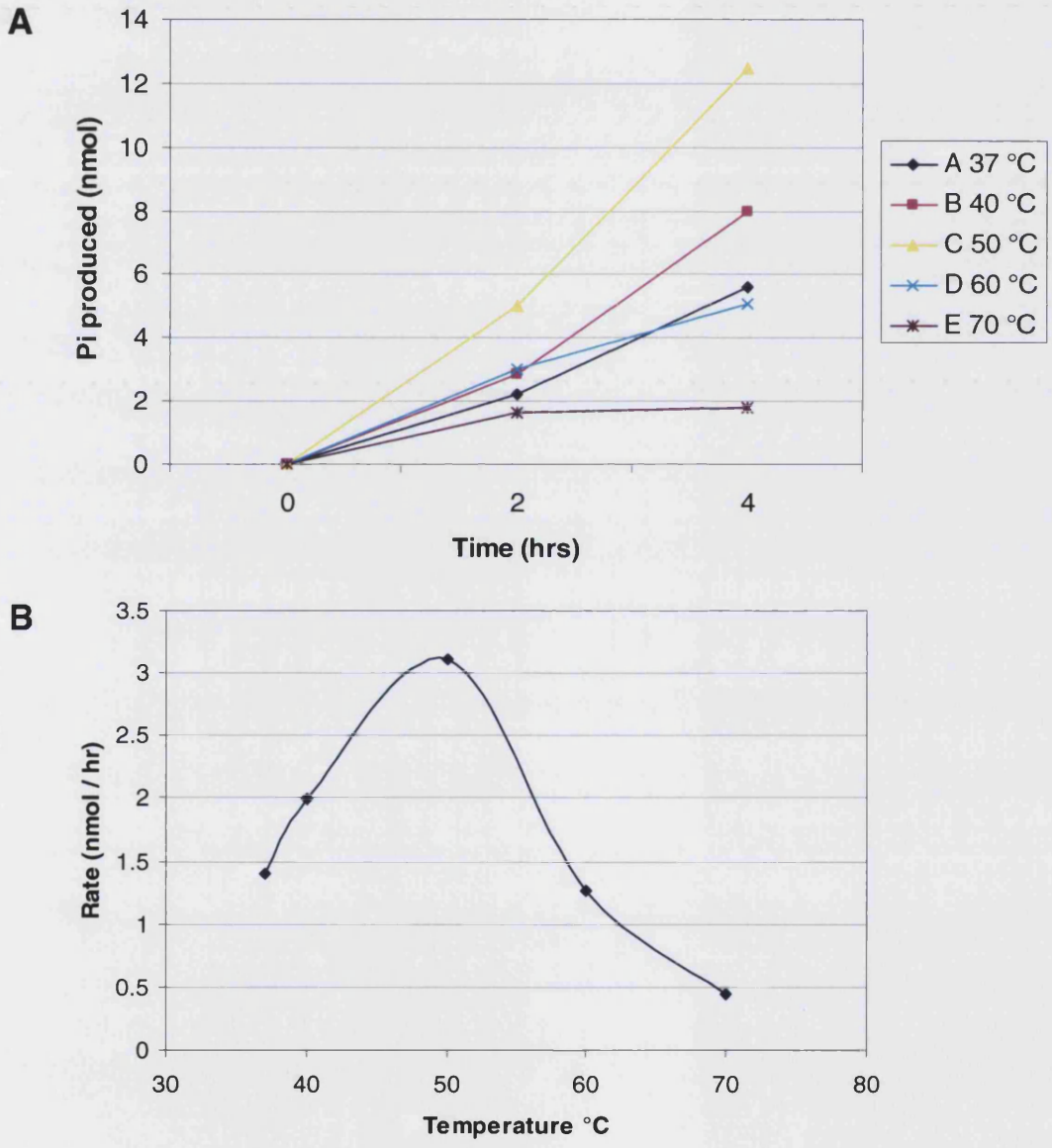


Figure 6-13 Effect of varying the reaction temperature. Mth203 reactions were performed at different temperatures and 50 µl samples assayed for Pi at 2 hr intervals (A). The rate of each reaction was estimated by calculating the regression line and was plotted against temperature (B).

Reactions were set up with four different concentrations of rATP substrate and assayed for Pi every half an hour (Fig. 6-14, next page). The reaction rates were near linear over the course of the experiment. The average rate of each reaction was estimated from its regression line. These values were fitted to the Hanes transformation of the Michaelis-Menten equation. The following parameters were obtained:

K_m = 1.131 mM rATP

V_{max} = 14.96 nmol / hr

K_{cat} = 13.85 min^{-1}

Hanes correlation coefficient = 0.75

A K_m of 1.131 mM denotes a very weak affinity for ATP, even for a DEAD-box protein. For example mammalian eIF4A has, for a DEAD-box protein, a relatively high K_m of 440 μM ATP (Lorsch and Herschlag, 1998). Also, the K_{cat} value of 13.85 min^{-1} is very low, although not unreasonable; mammalian eIF4A has a K_{cat} of 3 min^{-1} (Lorsch and Herschlag, 1998). The values stated above for mammalian eIF4A were obtained in the absence of its stimulatory cofactors eIF4B and eIF4H. Perhaps the low *in vitro* activity of Mth203 is an indication that it too requires other protein cofactors *in vivo*.

Only four data points were used in the Hanes transformation and this is reflected by the low correlation coefficient (0.75). Clearly, then, these parameters are not definitive. Another issue is the range of rATP concentrations that were tested. A limitation of the malachite green assay is that the nucleotide concentration must be kept below 300 μM due to background absorbance. Ideally, the highest concentration of substrate tested should be ten times the enzyme's K_m. With an estimated K_m of 1.1 mM for Mth203, for this assay that was impossible. This approach also assumes that the enzyme obeys Michaelis-Menten kinetics, which Mth203 may not.

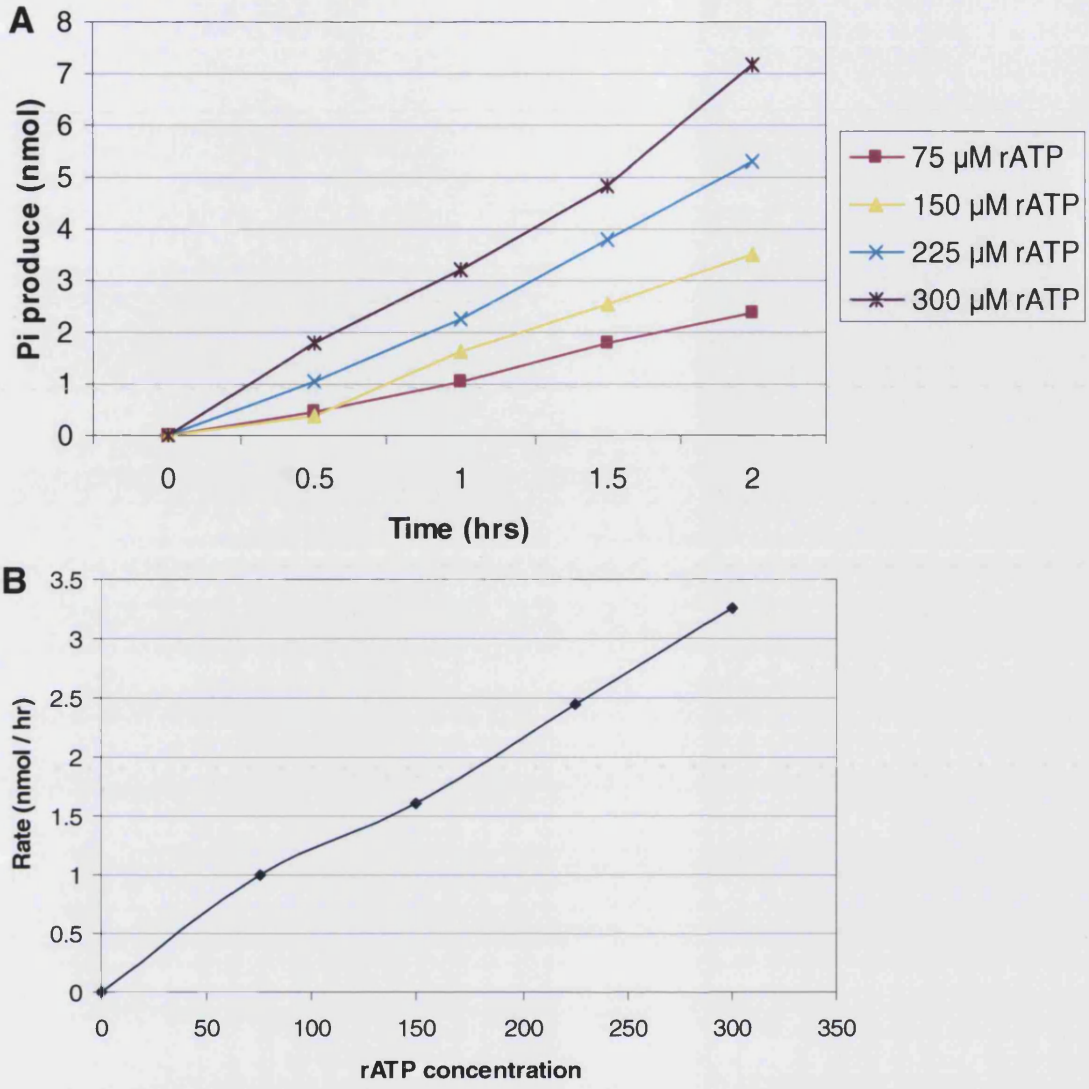


Figure 6-14 Effect of varying rATP concentration to determine kinetic parameters. Reactions containing different concentrations of rATP were set up. 50 µl samples were taken every half hr and assayed for Pi (A). The rate of each reaction was estimated by calculating the regression line and plotted against rATP concentration (B).

Mth203 was capable of hydrolysing any of the NTPs tested (Fig. 6-15). Although, from this data, it is unclear which one was the most efficient. This suggests Mth203 is promiscuous in its use of NTP substrate. This is usually considered to be more a characteristic of DEAH- than DEAD-box proteins. However, many DEAD-box proteins have been shown to use also dATP *in vitro* (Cordin *et al.*, 2006).

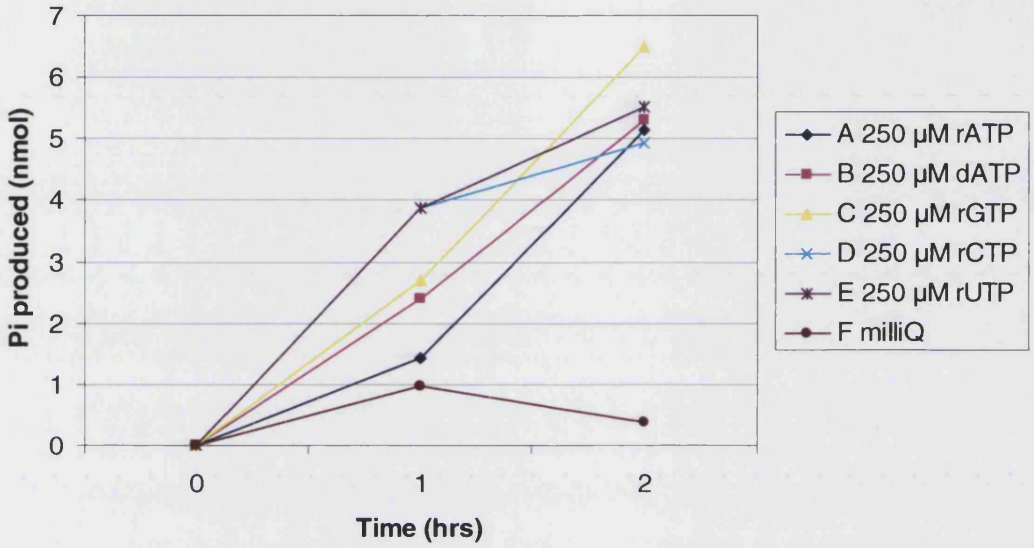


Figure 6-15 Effect of substituting different rNTPs or dNTPs. Reactions contained a 250 µM concentration of various nucleotide triphosphates (indicated). Data from a single experiment (not replicated).

Many DEAD-box proteins have been shown to possess a RNA dependent ATPase activity *in vitro*. For example Has1p, a DEAD-box protein believed to be involved in *S. cerevisiae* rRNA processing, is strictly dependent on the presence of RNA. RNA and DNA were added to ATPase reactions to investigate their effect on the ATPase activity (Fig. 6-16). The effect of adding RNA was ambiguous but adding DNA caused a small yet consistent increase in ATPase activity.

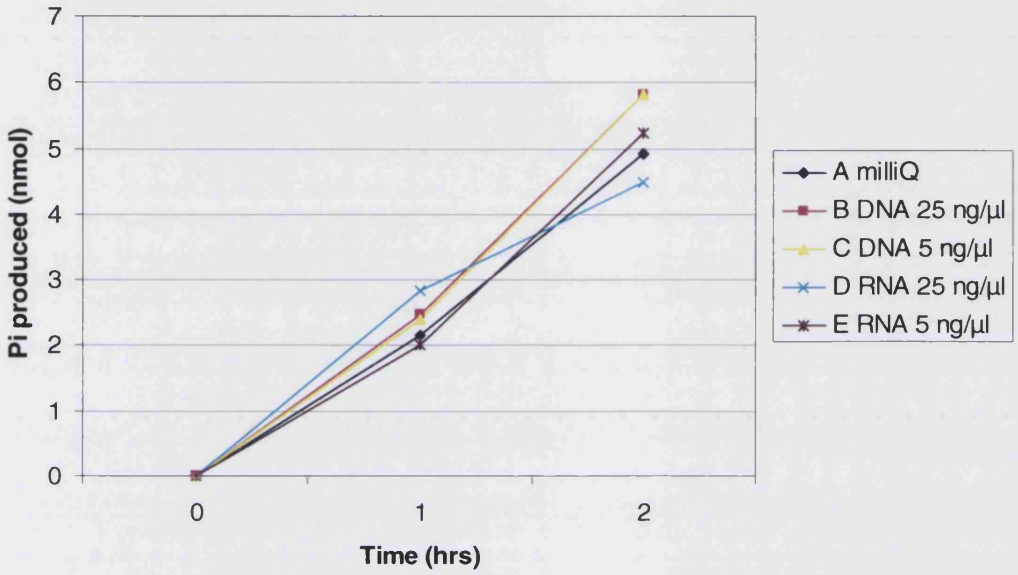


Figure 6-16 Effect of adding RNA or DNA. Mth203 reactions were supplemented with either total RNA from *S. cerevisiae* or high molecular weight DNA from salmon testes. Although neither had a considerable effect on ATPase activity, the DNA did give rise to a slight increase. Data from a single experiment (not replicated).

Oligonucleotide DNA was added to ATPase reactions in equimolar (360 nM), five times smaller and five times greater concentrations than Mth203 to investigate the effect on ATPase activity (Fig. 6-17). Oligonucleotide differs from salmon testes DNA in at least two important ways: it is single stranded and it gives far higher concentrations of free ends. Adding oligonucleotide caused a very small, but dose dependent increase in Mth203 ATPase activity. The fact that ATPase activity is only marginally stimulated by the addition of nucleic acid substrates may suggest a requirement for an additional protein cofactor(s), possibly Cdc6-1.

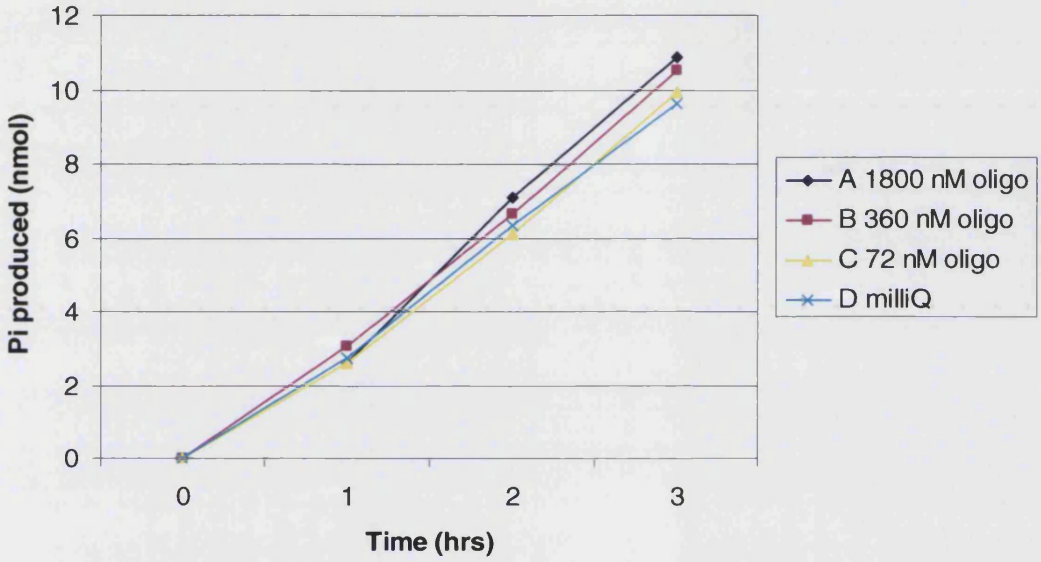


Figure 6-17 Effect of adding oligonucleotide. Mth203 reactions were supplemented with single stranded oligonucleotide (36mer). This produced a marginal concentration dependent increase in ATPase activity. Data from a single experiment (not replicated).

6.4 Discussion

The aim of this chapter was to investigate the function of Mth203. Mth203 has a series of nine sequence motifs that are characteristic of DEAD-box SF2 helicases (Fig. 6-1). There are three slight variations from the motif consensus sequences, but the protein may well still be functional (Table 6-1). SF2 motifs are found in numerous proteins involved in a wide range of processes. Therefore, their presence cannot be used to predict the function of Mth203 beyond basic activities such as NTP hydrolysis and polynucleotide translocation / unwinding, neither can the results of

database similarity searches, based on Mth203's primary sequence and predicted tertiary structure (Tables 6-2 and 6-3). Therefore, Mth203 was expressed and purified in order to begin an investigation into the protein's biochemical activities. Mth203 forms a dimer in solution. Dimer formation has been observed for only one other helicase to date, a DEAD-box protein from *M. jannashii* called Mj0669 (Story *et al.*, 2001). This may suggest a mechanistic or functional similarity between Mth203 and Mj0669. Unfortunately, the functional properties of Mj0669 have not yet been characterised, although it was noted by Story and colleagues (2001) that dimer formation is consistent with both of the two main models of helicase action (rolling and inchworm). Mth203 exhibits a measurable and repeatable ATPase activity. The results showed a very weak affinity for ATP and low rate of turnover. However, this is quite a common outcome of investigations into SF2 helicases and is thought to be due to a requirement for other protein(s).

7 General Discussion

Every organism must control the timing of initiation of DNA replication within its cell cycle. The systems that Bacteria and Eukarya use to achieve this have been extensively investigated, although much detail is still unknown. The archaeal system, on the other hand, remains almost completely unknown. Only a few molecular details are known, of which one of the more prominent has been the identification of an origin of replication in several archaeal species (Myllykallio *et al.*, 2000, Berquist and DasSarma, 2003). At first, this suggested a bacterial-like origin and therefore initiation system, however multiple origins have now been discovered in *S. solfataricus* and *Sulfolobus acidocaldarius* (Robinson *et al.*, 2004, Lundgren *et al.*, 2004), which suggests more affinity with eukaryotes. In addition, clear homologues of the eukaryotic initiation proteins, MCM and Cdc6, have been identified in almost every archaeal species for which the genome sequence is known, further strengthening the eukaryotic similarities. However, most other important components found in eukaryotes appear to be missing in the Archaea. Therefore, the aim of this project was to gain a better understanding of the way archaea control the initiation of DNA replication by identifying proteins that interact with *oriC*, MCM and Cdc6. This approach was based on several premises. First, a system for controlling initiation exists in archaea. Second, the *oriC* element and the MCM and Cdc6 initiation proteins are involved in this system. And lastly, other, as yet unidentified, components of this system interact directly with *oriC*, MCM or Cdc6. Therefore, the yeast one-hybrid system was used to screen a *M. thermautotrophicus* genomic library for proteins that interact with the *oriC* from that species. In addition, the yeast two-hybrid system was used to screen the same library for proteins that interacted with MCM and Cdc6-1 from *M. thermautotrophicus*.

7.1 What does oriC interact with?

All organisms initiate DNA replication from discrete sites on the chromosome known as origins. The position of each origin is marked by the binding of an initiator protein to a specific DNA sequence. In Eukarya, replication is initiated from multiple origin sites and the initiator is in fact a six-subunit assembly called the origin recognition complex (ORC) (Stillman, 2005). In contrast, replication in

Bacteria is controlled by a single initiator protein (DnaA) (Cunningham and Berger, 2005), which binds to repeated sequence elements called DnaA boxes within a single chromosomal origin (*oriC*). In archaea the situation is less clear, although similarities to both bacterial and eukaryotic systems have been noted (Robinson and Bell, 2005). For instance, the putative archaeal initiator has homology to eukaryotic Cdc6 and Orc1 proteins, as revealed by primary and tertiary structure similarities (Robinson *et al.*, 2004). However, the archaeal Cdc6/Orc1 is also structurally similar to bacterial DnaA, with an N-terminal AAA^+ domain and a C-terminal WH domain that contains a HTH DNA binding motif (Erzberger *et al.*, 2002). The archaeal Cdc6/Orc1 proteins can be divided into two clades (1 and 2) (Berquist and DasSarma, 2003). *M. thermautotrophicus* has one Cdc6/Orc1 protein from each clade, known as Cdc6-1 and Cdc6-2. There are also examples of non-initiator origin binding proteins. For example: in bacteria SeqA binds to and sequesters *oriC* after initiation has occurred in order to prevent re-replication (Riber and Lobner-Olesen, 2005) and in *S. cerevisiae* Abf1p acts as an enhancer of initiation (Raychaudhuri *et al.*, 1997).

Cdc6-1 is the best candidate for the initiator protein in *M. thermautotrophicus*. The gene for Cdc6-1 is located near the *M. thermautotrophicus* origin, an arrangement also observed for initiator genes and origins in many bacteria. In order to identify any further proteins that might bind to archaeal origins (e.g., homologues or analogues of SeqA or Abf1p), the *M. thermautotrophicus* genome was screened for origin interacting proteins (Chapter 3). Multiple independent fragments of Cdc6-1 were consistently identified by this screen. A control using a *M. thermautotrophicus* rDNA sequence demonstrated that these Cdc6-1 fragments interacted specifically with the origin element, as did the full-length Cdc6-1 protein. Interestingly, no other proteins were found to interact with the origin in this assay, despite the library being a 43 times coverage of the genome. Previous studies have shown that *M. thermautotrophicus* Cdc6-1 specifically recognises conserved 13 bp repeats that reside in the origin region (Capaldi and Berger, 2004, Kasiviswanathan *et al.*, 2006). Orthologues of Cdc6-1 have also been shown to bind origin sequences in other archaeal species including: *P. abyssii* (Matsunaga *et al.*, 2001), *S. solfataricus* (Robinson *et al.*, 2004) and *Archaeoglobus fulgidus* (Grainge *et al.*, 2003). It was discovered that the 13 bp *M. thermautotrophicus* Cdc6-1 binding site constitutes a

mini-ORB, the same sequence that is recognised by the *S. solfataricus* Cdc6-1 protein (Capaldi and Berger, 2004). Moreover, the binding of archaeal initiator proteins to short origin sequence repeats is reminiscent of the interaction between DnaA and DnaA boxes in bacteria. In contrast to Cdc6-1, *M. thermautotrophicus* Cdc6-2 binds weakly to dsDNA and has no preference for origin sequence repeats (Capaldi and Berger, 2004, Kasiviswanathan *et al.*, 2006). It is therefore worth noting that Cdc6-2 was not identified by the *oriC* interaction screen (Chapter 3). The above observations are consistent with the hypothesis that *M. thermautotrophicus* Cdc6-1 and Cdc6-2 have different functions, with Cdc6-1 being responsible for binding the origin. The Cdc6-1 fragments identified in the *oriC* screen all contained, as a minimum, the WH domain, which suggests that this domain was required for *oriC* interaction. Furthermore, analysis of Cdc6-1 mutants has demonstrated that the principal residues involved in this interaction map to a signature motif within the HTH region of the protein (Capaldi and Berger, 2004). Interestingly, this signature motif is conserved in the Cdc6-1 clade but absent from Cdc6-2 proteins. This distinction may explain the different binding affinities of *M. thermautotrophicus* Cdc6-1 and Cdc6-2 for the 13 bp origin repeats. It has also been shown that Cdc6-1 binds to the *M. thermautotrophicus* origin cooperatively (Capaldi and Berger, 2004). Interestingly, this is also a property of DnaA when it binds to the *E. coli oriC*. The above findings illustrate certain mechanistic parallels between the initiation of DNA replication in archaea and that of bacteria, i.e. both domains utilize a single initiator protein that binds cooperatively to short origin repeat sequences via a HTH motif.

Apart from Cdc6-1, no other proteins were identified by the *oriC* interaction screen. This suggests that *M. thermautotrophicus* encodes only one protein that interacts specifically with the origin, namely Cdc6-1. If this is indeed an accurate description of the situation in *M. thermautotrophicus* then it has, in itself, some important implications. Firstly, it means any proteins that need to be targeted specifically to the origin must be recruited there by Cdc6-1. That makes Cdc6-1 a truly pivotal component in the initiation process of *M. thermautotrophicus*. Furthermore, it means that *M. thermautotrophicus* does not form a multi-subunit complex at the origin akin to the ORC, a six-subunit assembly that binds the DNA at eukaryotic origins. Rather, the observation supports an analogy with bacteria in which a single initiator protein binds cooperatively to multiple binding sites in the *oriC*. This is in contrast

to the situation in *S. solfataricus* in which all three Cdc6/Orc1 proteins bind to the origins of replication. (Cdc6-1 and Cdc6-3 are thought to promote initiation whereas Cdc6-2 is believed to inhibit it.) Perhaps this disparity represents a fundamental difference between the initiation systems of the Euryarchaeota (to which *M. thermautotrophicus* belongs) and the Crenarchaeota (to which *S. solfataricus* belongs). This observation also suggests that *M. thermautotrophicus* does not prevent re-replication by sequestering its origin to a protein like SeqA. There must be another mechanism to prevent re-replication, possibly proteolysis of Cdc6-1, as occurs for eukaryotic Cdc6 (Sanchez *et al.*, 1999).

7.2 What does Cdc6-1 interact with?

M. thermautotrophicus Cdc6-1 interacted with itself in a yeast two-hybrid assay (Chapter 4). This is consistent with the observation that Cdc6-1 binds cooperatively to 13 bp mini-ORB repeats *in vitro* (Capaldi and Berger, 2004). The cooperative binding of DnaA is due to the self-association of AAA^+ domains between initiator molecules bound to *oriC* (Simmons *et al.*, 2003). Given the high degree of structural similarity between the AAA^+ domains of DnaA and Cdc6/Orc1 (Erzberger *et al.*, 2002), it has been suggested that this region might likewise be responsible for the cooperative binding of archaeal initiators (Capaldi and Berger, 2004). The Cdc6-1 library clone identified during screening encoded residues 178-383 of the full-length protein suggesting that this portion contains the region responsible for self-interaction.

M. thermautotrophicus MCM interacted with Cdc6-1 by yeast two-hybrid assay and by affinity co-purification (Chapters 4 and 5). This observation is in agreement with the results of previous studies in *M. thermautotrophicus* (Shin *et al.*, 2003, Kasiviswanathan *et al.*, 2005). An interaction has also been demonstrated between the homologues of MCM and Cdc6-1 in *S. solfataricus* (De Felice *et al.*, 2004b) and *P. aerophilum* (Kasiviswanathan *et al.*, 2005). Presumably, the purpose of this interaction is simply to recruit MCM to the origin just as DnaA recruits DnaB in bacteria. However, the association between the two proteins also appears to regulate their biochemical properties. MCM stimulates Cdc6-1's autophosphorylation activity (Kasiviswanathan *et al.*, 2005). MCM also substantially reduces Cdc6-1's

DNA binding affinity (Kasiviswanathan *et al.*, 2006). Is there a connection between these two activities? It has not been demonstrated, but perhaps the phosphorylation of Cdc6-1 causes it to have a lower affinity for origin DNA. This is interesting because it would suggest a possible mechanism by which origin-bound Cdc6-1 is inactivated upon MCM loading (an indicator of origin firing). In which case, such a mechanism would perform a function analogous to that of the regulatory inactivation of DnaA system in bacteria. It also suggests a parallel with eukaryotes in which Orc1 is phosphorylated and degraded by ubiquitin mediated proteolysis during the G1 to S-phase transition (Mendez *et al.*, 2002).

M. thermautotrophicus MCM did not interact with Cdc6-2 in a yeast two-hybrid assay (Chapter 4). This is in agreement with a two-hybrid result from a study conducted by Kasiviswanathan and colleagues (2005). However, in that same study it was shown that MCM and Cdc6-2 do interact by affinity co-purification. Full-length Cdc6-2 showed MCM binding but the WH domain alone did not (in contrast to Cdc6-1) suggesting that the N-terminal AAA^+ domain is also involved. Several further observations support a role for Cdc6-2 in loading the MCM helicase onto the origin. Firstly, it has sequence similarity to eukaryotic Cdc6, which performs this function in eukaryotes. It inhibits MCM just as DnaC, the helicase loader in bacteria, inhibits DnaB. Lastly, Cdc6-2 has no preference for binding origin repeats over random sequences and so is unlikely to be responsible for origin recognition.

A yeast two-hybrid screen of a *M. thermautotrophicus* genomic library was undertaken in order to identify novel interactions with Cdc6-1. 157 of the positives were analysed by restriction mapping giving 40 non-redundant clones. Next, non-specific interactions were eliminated by discarding clones that activated the reporter genes even when the Cdc6-1 bait was absent. That left 14 clones, which were then identified by sequence analysis and comparison with the published *M. thermautotrophicus* genome (Table 7-1)

Table 7-1 Positive library clones that interacted specifically with Cdc6-1.

ID of Positive	Mth ORF Code	Mth Annotation*
A11	203	ATP-dependent RNA helicase, eIF-4A family
A14	203	ATP-dependent RNA helicase, eIF-4A family
A16	1408	cobalamin biosynthesis protein G
A17	446	sensory transduction regulatory protein
A26	1907	conserved hypothetical protein
A34	151	methyl coenzyme M reductase system, A2 homolog
A42	492	ATP-dependent RNA helicase homolog
A48	1412	Cdc6-1
A67	1770	MCM
A94	1624	DNA topoisomerase I
A156	1458	hypothetical protein
A168	1215	fibrillarin-like pre-rRNA processing protein
A216	1715	phycocyanin alpha phycocyanobilin lyase CpcE related
A217	656	ATP-dependent RNA helicase related protein

*from *M. thermautotrophicus* complete genome gi15678031 (Smith *et al.*, 1997)

The results suggest that 13 different *M. thermautotrophicus* proteins interacted specifically with Cdc6-1. However, it seems unlikely that all 13 of these interactions are genuine and biologically meaningful. In order to have confidence in an interaction, one must first show that the full-length protein interacts with Cdc6-1 and second confirm the interaction using an independent method (e.g., affinity co-purification). Lack of time prohibited performing these analyses on all 13 proteins. Thus only the ones considered most interesting and plausible were taken further. Upon looking at the list of Cdc6-1 interactors, the most striking observation was that three of the genes are annotated "ATP-dependent RNA helicases" (Mth203, Mth492 and Mth656). Therefore, these proteins were chosen for further investigation. However, none of the interactions identified are totally implausible. For instance DNA topoisomerase I may be recruited to the origin by Cdc6-1 to relax the negative super-coils created by the replication forks after initiation has occurred. The interaction with methyl coenzyme M reductase may provide a means of coupling the timing of initiation to the level of methanogenesis and concomitant energy production as a kind of check-point mechanism. Various functions may be imagined for the conserved hypothetical proteins that were identified. The ease with which

plausible biological scenarios can be envisaged for the proteins interacting with Cdc6-1 supports the case for their specific and detailed investigation in the future.

Mth203, Mth492 and Mth656 are all members of the superfamily 2 (SF2) class of DNA and RNA helicases. It is well known that SF2 proteins often have key functions in DNA and RNA processing systems (Singleton and Wigley, 2002). Therefore, Mth203, Mth492 and Mth656 made good candidates for involvement in archaeal replication licensing, replication or cell-cycle control and so were selected for further investigation. As mentioned above, the next step was to test interaction between each of the three full-length proteins and Cdc6-1. This was necessary to establish whether a fragment interaction was a valid representation of an interaction with the full-length protein or due to an artifactual binding site exposed by removal of endogenous domains.

Full-length Mth203 protein interacted with Cdc6-1, whereas full-length Mth492 and Mth656 did not. The Mth492 and Mth656 library fragments both essentially represent domain 1 of the SF2 core fold of their respective full-length proteins. This suggests that the tertiary structure elements are recognised by Cdc6-1 in an equivalent manner in both proteins. Why is no binding observed in the context of full-length proteins? There are three possible explanations for this. A) An artifactual Cdc6-1 binding site was exposed on each of the library fragments, which would not normally be available *in vivo*. B) The domain 1 binding site would normally be available *in vivo* but is obstructed by the GAL4 DNA BD and only exposed by removal of domain 2 and the C-terminal extension (as in the library fragments). C) The two-hybrid screen identified genuine interactions but the full-length proteins didn't interact due to inappropriate physiological conditions (temperature, pH, etc.) for these proteins. These hypotheses could be tested by using a different method for detecting protein-protein interactions that does not involve a bulky N-terminal fusion (e.g., affinity co-purification).

The most important finding in Chapter 4 was the discovery of a novel interaction between the *M. thermautotrophicus* proteins Cdc6-1 and Mth203. Two independent Mth203 fragments (11 and 14) were identified in the Cdc6-1 library screen. Both fragments and the full-length Mth203 protein interacted specifically with Cdc6-1 in a

two-hybrid assay. Mth203 has a ~50 residue C-terminal extension with respect to the SF2 helicase core. This extension was present in both fragments. Furthermore, fragment 11 contained just the 53 C-terminal residues of Mth203, indicating that this extension alone was sufficient for Cdc6-1 binding. As neither fragment contained Mth203's domain 1, the mode of Cdc6-1 binding appears to be different from that of Mth492 and Mth656. Moreover, as the extension is in a non-conserved portion of the protein, this particular mode of Cdc6-1 binding is likely to be specific to Mth203.

The interaction between full-length Mth203 and Cdc6-1 was also confirmed by affinity co-purification (Chapter 5). Affinity co-purification is claimed to be very sensitive, capable of detecting interactions with dissociation constants (K_ds) in the μM range (Ausubel, 1987). The two-hybrid system is more sensitive still. Therefore, the current results suggest that the interaction between Cdc6-1 and Mth203 has an affinity with a K_d of at most 1 μM. In order to obtain a more precise measurement of the strength of interaction between Cdc6-1 and Mth203, a different method must be used, such as surface plasmon resonance. The two-hybrid system is capable of detecting even very transient interactions. Affinity co-purification on the other hand, can only detect relatively stable interactions because they must be able to persist throughout the binding and washing steps. Detection of Cdc6-1/Mth203 binding by affinity co-purification therefore suggests that it is a relatively stable interaction. This observation is consistent with the formation of a complex or recruitment of the protein to some location. As with binding affinity, the kinetic parameters of this interaction may be measured by using surface plasmon resonance.

7.3 What does Mth203 do?

What biochemical activities does Mth203 have and what is its function in the cell? The gene possesses the proper transcription and translation elements (TATA-box, Shine-Dalgarno site, etc.), which suggests that it is expressed *in vivo*. The primary sequence exhibits all nine intact DEAD-box SF2 helicase motifs, indicating Mth203 is a functional protein. However, these motifs cannot be used to infer substrate specificity or other catalytic characteristics. Therefore, methods were developed to purify recombinantly expressed His-tagged Mth203 in order to begin functional studies in earnest. Using size-exclusion chromatography, Mth203 was observed to

form a dimer. Is dimer formation a requisite for its function? Mth203 exhibited a measurable NTPase activity suggesting that it is catalytically functional. This activity required Mg^{2+} or Mn^{2+} and had an optimum temperature of approximately 50°C. The enzyme had a very weak affinity for ATP and a low rate of turnover. However, these observations only reflect Mth203's intrinsic activity and were made using an *in vitro* assay. Therefore, this low activity may imply a requirement for another protein (e.g., Cdc6-1) or a particular polynucleotide substrate, although the activity was not significantly stimulated by bulk DNA or RNA.

The aim of the work presented in Chapter 6 was to investigate Mth203's function. Time ran out before that function could be narrowed down very far. However, it was demonstrated that Mth203 contained intact SF2 helicase motifs and the recombinantly expressed protein exhibited NTPase activity *in vitro*. Given the broadest definition of a SF2 helicase, these observations suggest that Mth203 is a functional protein and uses energy from the hydrolysis of NTPs to elicit some mechanical event on DNA or RNA. This provides a basis for speculation on the biological function of Mth203.

Two key activities of SF2 helicases are: the translocation along, and the unwinding of polynucleotides. But what if Mth203 had neither of these properties? Clearly this would be unusual for a protein with SF2 motifs and would raise the question: what is the energy from ATP hydrolysis being used for? However, an example of just such a protein is *E. coli* SecA (de Keyzer *et al.*, 2003). The role of SecA is to provide the driving force for the secretion of polypeptides across the cytoplasmic membrane. So, although unusual, a SF2 protein with a function unrelated to polynucleotide translocation / unwinding is not unprecedented. It is also quite possible that Mth203 has a completely novel function for a SF2 protein.

Suppose that Mth203's function did depend on polynucleotide translocation but <u>not</u> on unwinding. Would it translocate on DNA or RNA? With what degree of processivity? And on which strand, 5'→3' or 3'→5'? There are several examples of SF2 proteins that translocate on DNA when the objective is not strand separation. For instance, UvrB is a component of the nucleotide excision repair pathway in prokaryotes (Theis *et al.*, 2000). It tracks along DNA, together with UvrA, in an

ATP-dependent manner until it encounters a lesion. RecG's function is to rescue stalled replication forks by regressing them to a point at which one of the DNA damage bypass pathways may act (Sharples *et al.*, 1999). Eukaryotic Snf2 family proteins remodel chromatin by tracking along DNA (Thoma *et al.*, 2005). ATP hydrolysis is coupled to DNA translocation by SF2 motifs in the R-subunit of all Type 1 restriction modification enzymes (Singleton and Wigley, 2002). Strand separation activity is not required for this translocation activity (Stanley *et al.*, 2006). Therefore, the presence of SF2 motifs does not necessarily imply strand separation functionality. The examples outlined above show that the SF2 protein core provides the ATPase motor that allows various holoenzymes to translocate along polynucleotides. Additional domains, which may vary from one protein to another, are responsible for modulating DNA unwinding, mediating protein-protein interactions or determining substrate specificity.

Now suppose that Mth203s function involves translocation and unwinding of a polynucleotide substrate. There are many SF2 proteins with these properties. Would Mth203 unwind dsDNA, dsRNA or a DNA/RNA hybrid? Would it have a preference for a particular type of secondary structure (such as flaps, forks, bubbles or three-/four-way junctions)? Or would it have a preference for a particular base sequence? There are several examples of SF2 helicases that unwind DNA. MPH1 is a 3'→5' DNA helicase that is involved in homologous recombination mediated mutation avoidance in *S. cerevisiae* (Prakash *et al.*, 2005). The ubiquitous RecQ family of SF2 DNA helicases has an unusually wide breadth of activities including roles in DNA replication, recombination and repair (Bennett and Keck, 2004). There are numerous examples of SF2 proteins that unwind RNA and they have roles in various aspects of RNA metabolism (Rocak and Linder, 2004, Cordin *et al.*, 2006). eIF4A, the prototypical DEAD-box protein, unwinds the 5' untranslated region of eukaryotic mRNAs preparing them for translation (Rocak and Linder, 2004). And in *S. cerevisiae*, 15 DEAD-box proteins (out of 26) are needed to unwind short regions of dsRNA involved in ribosome biogenesis (Rocak and Linder, 2004).

SF2 proteins clearly have a broad range of very diverse roles. It is also worth noting that SF2 proteins often fulfil these roles by acting as part of a complex (e.g., SecA as part of SecYEG, UvrB as part of UvrABC and Snf2 as part of RSC) and that their

function is heavily influenced by the proteins with which they interact. Therefore, it is likely that Mth203 also acts as part of a complex. This possibility raises the questions: what does that complex do? What part does Mth203 play? Which other proteins participate? A clue to the answers to these questions may come from the observation that Mth203 interacts physically with Cdc6-1 (Chapters 4 & 5) and this premise is discussed further below. Clearly, though, further work is necessary to glean a better insight into the activities of Mth203.

7.4 Why do Cdc6-1 and Mth203 interact?

The results presented in Chapter 6 demonstrate that Mth203 is a functional NTPase and possesses motifs characteristic of a SF2 helicase. Given the activities of characterised SF2 proteins, these observations suggest that Mth203 exhibits polynucleotide translocation and possibly also strand unwinding. Mth203 also interacts directly with Cdc6-1. The null hypothesis is that this interaction has no biological function. However, this interaction is specific, stable and involves a defined region of Mth203. Therefore, it is a reasonable hypothesis that this interaction does indeed elicit some biological function. What is the purpose of that function?

Cdc6-1 is the putative initiator protein and it interacts specifically with the *M. thermautotrophicus* origin. Therefore, an important consideration is whether the Cdc6-1/Mth203 interaction occurs at the origin or away from it. First, suppose that Cdc6-1 recruits Mth203 to perform some function at the origin. This would seem to suggest that Mth203 acts on DNA and not RNA. What effect does Mth203 have on DNA at the origin? Mth203 may unwind the origin DNA strands directly in order to form the open complex in preparation for MCM loading and replisome assembly. In bacteria, open complex formation is brought about by cooperative binding of many ATP-charged DnaA molecules to binding sites at *oriC*. In eukaryotes, the open complex is thought to be formed when the bound MCM helicase is activated by Cdc7-Dbf4 kinase. In *M. thermautotrophicus*, Cdc6-1 might act in a DnaA-like manner to open the DNA at the origin. However, it is not yet known how archaea accomplish open complex formation, so it is also possible that this process could involve an additional protein such as Mth203. Alternatively, Mth203 may play a

more subtle role in open complex formation such as chromatin remodelling. Eukaryotic chromatin remodelling proteins such as SWI2/SNF2 contain sequence motifs characteristic of SF2 helicases (Fig. 7-1). They act as part of a complex and use energy derived from NTP hydrolysis to generate superhelical torsion and destabilize histone-DNA interactions. By doing this, they promote a chromatin structure that is conducive to transcription. Perhaps in *M. thermautotrophicus*, Mth203 acts at the origin to promote a chromatin structure that is conducive to initiation of DNA replication. Or, Mth203 may prepare the origin for initiation by unwinding secondary structures. Under supercoiling conditions such as those found *in vivo*, the inverted repeat sequences found in many archaeal origins may form secondary structures (e.g., hairpins) that could impede initiation. Mth203 could be recruited in order to unwind such structures.

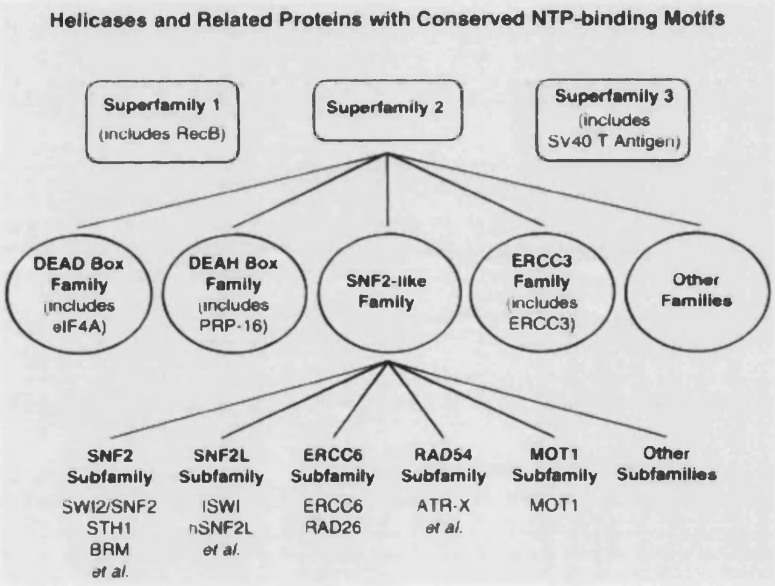


Figure 7-1 The SNF-2 like family of proteins contains motifs that are also present in SF2 helicases (Pazin and Kadonaga, 1997).

Perhaps Mth203 is involved not in initiation, but in DNA repair. Mth203 may be part of some repair system that Cdc6-1 recruits to ensure integrity of the origin before replication is initiated. In bacteria, a SF2 helicase called UvrB forms a complex with UvrA and the UvrAB complex tracks along chromosomal DNA until it encounters a lesion. Then UvrC is recruited to cleave the DNA and begin the repair process (Theis *et al.*, 2000). Perhaps Mth203 is recruited to the origin by Cdc6-1,

tracks along the DNA and if it encounters damage it halts DNA replication until the lesion is repaired, thereby providing a DNA damage checkpoint. Mfd is a bacterial SF2 helicase that functions in transcription coupled DNA repair (Roberts and Park, 2004). It releases transcription elongation complexes stalled by DNA damage or other obstacles. Mfd's function is thought to involve DNA translocation without the strand separation associated with classical helicases (Roberts and Park, 2004). Therefore, Mth203 could be recruited to the origin and translocate along chromosomal DNA in order to clear stalled RNA polymerases ahead of the advancing replication fork.

Mth203's role at the origin may be independent of any translocation or helicase activities that the enzyme might have. An important consideration about the interaction between Mth203 and Cdc6-1 is whether it modulates Cdc6-1's biochemical activities (origin DNA binding, MCM binding, ATPase and autophosphorylation). If so, then a regulatory function may be envisaged. In bacteria a homo-dimeric protein called Hda interacts with DnaA and inactivates it by stimulating its intrinsic ATPase activity in response to the loading of DNA polymerase III (a hallmark of replisome assembly) (Camara *et al.*, 2005). Perhaps Mth203's interaction may be involved in stimulating the inactivation of Cdc6-1 (by ATPase or autophosphorylation) in response to some initiation event in archaea. A comparison may also be made to the eukaryotes, in which Orc1 is phosphorylated and degraded by ubiquitin mediated proteolysis during the G1 to S-phase transition (Mendez *et al.*, 2002).

Now suppose that the interaction between Cdc6-1 and Mth203 occurs away from the origin. This would seem to suggest that the interaction is not directly involved in the process of initiating DNA replication. However, it could still be involved in regulating that process. For instance, perhaps Mth203 sequesters free Cdc6-1 in order to prevent it binding to the origin and functioning as the initiator. In which case, Mth203 might be a component of a checkpoint at the start of DNA replication or system to prevent over-initiation. Such a role would be reminiscent of that of geminin in higher eukaryotes. Geminin inhibits pre-replication complex formation during S- and G2-phases by sequestering Cdt1, a protein required for MCM loading (Saxena and Dutta, 2005). Such a role for Mth203 would also have some features in

common with Hda, a protein that binds to and down regulates the initiator in bacteria. An homologue of Mth203 called deaD is upregulated in response to cold shock in *Methanococcoides burtonii* (Lim *et al.*, 2000). Perhaps Mth203 delays the onset of DNA replication in response to stress in *M. thermautotrophicus*. But what about Mth203's predicted SF2 helicase activities? Perhaps Mth203 has a dual role: 1) helicase related and 2) Cdc6-1 sequestration.

Reversing the argument that Cdc6-1 recruits Mth203 highlights the possibility that Mth203 may in fact recruit Cdc6-1. If so, then perhaps Mth203 recruits Cdc6-1 for its ability to recruit MCM. For instance, Mth203 may be involved in transcription. The eukaryotic MCM complex has been shown to interact with RNA polymerase (Holland *et al.*, 2002). It has been suggested that MCM may unwind the DNA ahead of the RNA polymerase (Snyder *et al.*, 2005). Therefore, Mth203 may act as part of the transcription initiation complex, interacting with Cdc6-1 to recruit MCM to promoters.

Alternatively, Cdc6-1 may interact with Mth203 away from the origin in order to modulate Mth203's activities (NTPase, translocation or unwinding). Is this modulation stimulatory or inhibitory? Does Cdc6-1 modulate Mth203 by simple protein-protein interaction or by phosphorylation? Cdc6-1 has autophosphorylation activity, which is an intramolecular reaction. However, proteins that autophosphorylate are often also protein kinases. Perhaps Cdc6-1 phosphorylates Mth203. Either way, it may provide an interesting mechanism for regulating Mth203's activity according to the cell-cycle. The level of Cdc6-1 may rise and fall over the course of the cell cycle, as has been demonstrated in *S. solfataricus* (Robinson *et al.*, 2004). If Cdc6-1 modulates Mth203 then this may lead to a concurrent rise and fall in Mth203's activity. There are various possible roles for a SF2 helicase whose activity varies with the cell cycle. For instance Mth203 may be involved in DNA repair. An example of an enzyme, with SF2 motifs, that is involved in DNA repair is Rad54. Rad54's expression is cell cycle regulated in *S. cerevisiae*, being predominantly expressed in G1 (Spellman *et al.*, 1998). Perhaps Mth203's activity is stimulated by a high level of Cdc6-1 during G1 to repair chromosomal DNA in preparation for S-phase.

How cells control the timing of DNA replication within the cell cycle is an important issue in biology. The bacterial system is based on variations in the concentration of free initiator protein. In contrast, the eukaryotes have a more complex system that is characterised by rapid shifts in protein kinase activity, which serve as discrete switches. Very little is known about how the archaeal system works. However, some key components involved in the initiation of DNA replication have been identified (*oriC*, MCM and Cdc6/Orc1). The broad aim of this project was to shed some light on the molecular system that the Archaea use to control DNA replication. The main premise, on which the work was based, was that this system includes other proteins that interact directly with *oriC*, MCM or Cdc6/Orc1. The principle that the archaeal system is based upon is still not clear. Nonetheless, the work presented here does contribute to the understanding of archaeal DNA replication. Cdc6-1's importance in the initiation of DNA replication is promoted by the observation that it was the only protein to interact specifically with the origin. Furthermore, 13 proteins were identified by a screen conducted to find proteins that interact with Cdc6-1. Three of the proteins possessed motifs characteristic of SF2 helicases (Mth203, Mth492 and Mth656). Full-length Mth203 interacted with Cdc6-1 in two-hybrid and affinity co-purification assays. Recombinant Mth203 protein was expressed and purified. It exhibited a low intrinsic ATPase activity, but it may well be stimulated by other proteins or polynucleotide substrates *in vivo*. It is not yet clear whether Mth203 is involved in the control of DNA replication or some other related process. However, these findings provide a valuable starting point for further investigation.

8 Further Work

8.1 Interaction Studies

In Chapter 4 it was shown that the 53 residue C-terminal extension of Mth203 was sufficient for interaction with Cdc6-1. It would be interesting to test a Mth203 C-terminal deletion mutant for interaction with Cdc6-1 by two-hybrid assay to determine whether this extension is necessary for binding. Which portion of Cdc6-1 is responsible for the interaction with Mth203? This issue may be addressed by making constructs to encode different portions of Cdc6-1 (e.g., the AAA^+ domain and the WH domain) and then test them for interaction with Mth203 in a two-hybrid assay. It may also be informative to determine the stoiciometry of the Cdc6-1 interaction. According to the crystal structure, the C-terminus of Mj0669 is positioned well away from the dimer interface (Story *et al.*, 2001). If Mth203 dimerises in an equivalent manner to Mj0669 and binds to Cdc6-1 through its C-terminal extension, then each subunit may bind a single Cdc6-1. This hypothesis may be tested by combining the two proteins and measuring the size of the complex formed using a calibrated size-exclusion column. The Cdc6-1/Mth203 interaction analyses presented here were largely qualitative. Therefore, it would be useful to perform a quantitative analysis of the strength of the interaction (equilibrium association constant) and the rate at which association and dissociation occur (dissociation rate constant). Both of these parameters may be measured using surface plasmon resonance (Ausubel, 1987, Karlsson and Falt, 1997). Cdc6-1 and Mth203 have been shown to interact in a heterologous system (yeast two-hybrid) and *in vitro* (affinity co-purification). However, it would also be desirable to demonstrate that this interaction occurs *in vivo*. This may be accomplished by co-immunoprecipitation. An antibody would be used to precipitate Mth203 from a *M. thermautotrophicus* whole-cell extract. The precipitated proteins would then be separated by SDS-PAGE and Cdc6-1 identified by Western blotting or mass spectrometry. This method could also be used to identify any additional protein with which Mth203 interacts.

Mth203's function may be regulated by interactions with other proteins. Furthermore, Mth203 may function as part of a protein complex as many SF2

helicases are known to do. Therefore, it may be informative to screen the *M. thermautotrophicus* two-hybrid prey library using Mth203 as bait. As described above, Cdc6-2 is thought to load MCM onto the origin DNA, an important event that may be regulated by other proteins. Therefore, it may also be informative to screen the library with Cdc6-2 as bait. No specific positives were identified in the MCM screen (not even Cdc6-1) so it may be worthwhile to check everything and re-screen with MCM as bait.

Cdc6-1 recognises and binds to short repeat elements at the *M. thermautotrophicus* origin. However, there may be other Cdc6-1 binding sites elsewhere on the chromosome. For instance, there may be a cluster of high affinity Cdc6-1 binding sites, analogous to the *datA* locus in *E. coli* (Kato, 2005), which participates in a novel archaeal initiator titration mechanism. Therefore, it may be worthwhile to use the one-hybrid system to screen a *M. thermautotrophicus* genomic DNA bait library for DNA elements that interact with a Cdc6-1 protein prey. It may also be beneficial to use a filter binding assay to measure the affinity of Mth203 for various polynucleotide substrates such as: single or double stranded DNA, RNA and heteroduplex and various secondary structures (flaps, forks, Holliday junctions, etc.).

8.2 Expression Studies

As part of the characterisation of a novel protein such as Mth203, it is important to confirm that it is expressed *in vivo*. Reverse transcription quantitative PCR (qPCR) could be used to determine whether the *mth203* gene is transcribed. Then Western blotting with an antibody raised against Mth203 could be used to show that the transcript is also translated. qPCR and Western blotting could also be used to investigate further detail of Mth203's expression. Under what conditions is Mth203 expressed? Is it induced in response to stress? Is it regulated according to the cell-cycle? The latter hypothesis could be investigated by taking samples from a synchronised *M. thermautotrophicus* culture. The technique for establishing such a synchronised culture has been developed by Dr. James Chong (*pers. comm.*) and involves limiting the culture's hydrogen source. It is also possible that Mth203 may be post-translationally modified *in vivo*. This possibility may be investigated using 2D gel electrophoresis and mass spectrometry.

8.3 Biochemistry

Further Mth203 ATPase assays may be conducted in order to glean more insight into the protein's function. Many SF2 helicases function as part of multi-protein complexes. Mth203 may also act in a complex and require the other components for full ATPase activity. Therefore, it may be revealing to test the effect of other proteins on Mth203's activity. Given that Mth203 interacts stably with Cdc6-1, it makes sense for Cdc6-1 to be the first protein tested. But, since Cdc6-1 has an ATPase activity of its own, an ATPase defective mutant would need to be used so as not to confound the results. Additional proteins may also be tested for an effect on Mth203's activity, depending on results from the Mth203 interaction screens (two-hybrid and co-immunoprecipitation). SF2 helicase activities are often enhanced by the presence of particular forms of nucleic acids. This may give an indication of the type of substrate an enzyme acts upon *in vivo*. Therefore, various forms of polynucleotide substrates could be tested for their effect on Mth203's ATPase activity.

The presence of SF2 helicase motifs does not always imply strand separation functionality. Therefore, a helicase assay should be used to determine whether or not Mth203 has an intrinsic ability to unwind a polynucleotide duplex. The most widely accepted and direct assay of helicase activity is to start with a duplex made up of two single-stranded polynucleotides of different length (the shorter of which is radiolabelled), incubate it with the helicase and then the proportion of short poynucleotide that has been unwound is measured by electrophoresis and autoradiography. This assay could also be used to determine whether Mth203's helicase activity requires the presence of another protein(s) such as Cdc6-1 or any other proteins identified by a Mth203 interaction screen. It could also be used to determine whether Mth203 has any polynucleotide substrate specificity.

The crystal structures of Cdc6/Orc1 proteins from *P. aerophilum* and *A. pernix* have been solved. However, both these proteins are within the Cdc6-2 clade. Given the predicted functional differences of Cdc6-1 and Cdc6-2 from *M. thermautotrophicus* it may be interesting to determine the structure of Cdc6-1. Furthermore, it may be informative to determine Cdc6-1's structure as a complex with one of the 13 bp

repeat elements found in the *M. thermautotrophicus* origin. This may reveal whether Cdc6-1 has the ability to distort DNA in a manner similar to DnaA.

If an antibody against Mth203 were obtained, then immunolocalization may be conducted. Initially, light microscopy could be investigated but, for finer resolution, immuno-gold staining and transmission electron microscopy may be used. This may reveal whether Mth203 is associated with a particular cellular structure such as the nucleoid.

8.4 Genetics

Genetic systems for *M. thermautotrophicus* are currently being developed by Dr. James Chong at the University of York (*pers. comm.*). Therefore, genetic analysis of this organism may be possible in the near future. The first aim would be to make a Mth203 knock out mutant and investigate its phenotype. If Mth203 is a key component of DNA replication control then knocking it out would be expected to be lethal. However, if the mutant is viable then the effect on cell growth characteristics, morphology and susceptibility to DNA damage may be investigated. Secondly, Mth203 could be over-expressed and the effect of this on the *M. thermautotrophicus* phenotype investigated, as above.

9 References

APARICIO, O. M., WEINSTEIN, D. M. & BELL, S. P. (1997) Components and dynamics of DNA replication complexes in S. cerevisiae: redistribution of MCM proteins and Cdc45p during S phase. *Cell,* 91, 59-69.

AUSUBEL, F. M. (1987) *Current protocols in molecular biology,* New York, Wiley.

BARNS, S. M., DELWICHE, C. F., PALMER, J. D. & PACE, N. R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A,* 93, 9188-93.

BELL, S. P. & STILLMAN, B. (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature,* 357, 128-34.

BENNETT, R. J. & KECK, J. L. (2004) Structure and function of RecQ DNA helicases. *Crit Rev Biochem Mol Biol,* 39, 79-97.

BERNANDER, R. (2000) Chromosome replication, nucleoid segregation and cell division in archaea. *Trends Microbiol,* 8, 278-83.

BERNANDER, R. (2003) The archaeal cell cycle: current issues. *Mol Microbiol,* 48, 599-604.

BERNANDER, R. & POPLAWSKI, A. (1997) Cell cycle characteristics of thermophilic archaea. *J Bacteriol,* 179, 4963-9.

BERNSTEIN, D. A., ZITTEL, M. C. & KECK, J. L. (2003) High-resolution structure of the E.coli RecQ helicase catalytic core. *Embo J,* 22, 4910-21.

BERQUIST, B. R. & DASSARMA, S. (2003) An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, Halobacterium sp. strain NRC-1. *J Bacteriol,* 185, 5959-66.

BLOW, J. J. & HODGSON, B. (2002) Replication licensing--defining the proliferative state? *Trends Cell Biol,* 12, 72-8.

BOYE, E., STOKKE, T., KLECKNER, N. & SKARSTAD, K. (1996) Coordinating DNA replication initiation with cell growth: differential roles for DnaA and SeqA proteins. *Proc Natl Acad Sci U S A,* 93, 12206-11.

BULT, C. J., WHITE, O., OLSEN, G. J., ZHOU, L., FLEISCHMANN, R. D., SUTTON, G. G., BLAKE, J. A., FITZGERALD, L. M., CLAYTON, R. A., GOCAYNE, J. D., KERLAVAGE, A. R., DOUGHERTY, B. A., TOMB, J. F., ADAMS, M. D., REICH, C. I., OVERBEEK, R., KIRKNESS, E. F., WEINSTOCK, K. G., MERRICK, J. M., GLODEK, A., SCOTT, J. L., GEOGHAGEN, N. S. & VENTER, J. C. (1996) Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science,* 273, 1058-73.

CAMARA, J. E., BREIER, A. M., BRENDLER, T., AUSTIN, S., COZZARELLI, N. R. & CROOKE, E. (2005) Hda inactivation of DnaA is the predominant mechanism preventing hyperinitiation of Escherichia coli DNA replication. *EMBO Rep,* 6, 736-41.

CAPALDI, S. A. & BERGER, J. M. (2004) Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon Methanothermobacter thermoautotrophicus. *Nucleic Acids Res,* 32, 4821-32.

CARMEL, A. B. & MATTHEWS, B. W. (2004) Crystal structure of the BstDEAD N-terminal domain: a novel DEAD protein from Bacillus stearothermophilus. *Rna,* 10, 66-74.

CARUTHERS, J. M., JOHNSON, E. R. & MCKAY, D. B. (2000) Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc Natl Acad Sci U S A,* 97, 13080-5.

CHONG, J. P., HAYASHI, M. K., SIMON, M. N., XU, R. M. & STILLMAN, B. (2000) A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. *Proc Natl Acad Sci U S A,* 97, 1530-5.

COIN, F., MARINONI, J. C., RODOLFO, C., FRIBOURG, S., PEDRINI, A. M. & EGLY, J. M. (1998) Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. *Nat Genet,* 20, 184-8.

COOPER, S. & HELMSTETTER, C. E. (1968) Chromosome replication and the division cycle of Escherichia coli B/r. *J Mol Biol,* 31, 519-40.

CORDIN, O., BANROQUES, J., TANNER, N. K. & LINDER, P. (2006) The DEAD-box protein family of RNA helicases. *Gene,* 367, 17-37.

CUNNINGHAM, E. L. & BERGER, J. M. (2005) Unraveling the early steps of prokaryotic replication. *Curr Opin Struct Biol,* 15, 68-76.

DE FELICE, M., ESPOSITO, L., PUCCI, B., DE FALCO, M., MANCO, G., ROSSI, M. & PISANI, F. M. (2004a) Modular organization of a Cdc6-like protein from the crenarchaeon Sulfolobus solfataricus. *Biochem J,* 381, 645-53.

DE FELICE, M., ESPOSITO, L., PUCCI, B., DE FALCO, M., ROSSI, M. & PISANI, F. M. (2004b) A CDC6-like factor from the archaea Sulfolobus solfataricus promotes binding of the mini-chromosome maintenance complex to DNA. *J Biol Chem,* 279, 43008-12.

DE KEYZER, J., VAN DER DOES, C. & DRIESSEN, A. J. (2003) The bacterial translocase: a dynamic protein channel complex. *Cell Mol Life Sci,* 60, 2034-52.

DELONG, E. F., WU, K. Y., PREZELIN, B. B. & JOVINE, R. V. (1994) High abundance of Archaea in Antarctic marine picoplankton. *Nature,* 371, 695-7.

DIONNE, I. & BELL, S. D. (2005) Characterization of an archaeal family 4 uracil DNA glycosylase and its interaction with PCNA and chromatin proteins. *Biochem J,* 387, 859-63.

DIONNE, I., NOOKALA, R. K., JACKSON, S. P., DOHERTY, A. J. & BELL, S. D. (2003) A heterotrimeric PCNA in the hyperthermophilic archaeon Sulfolobus solfataricus. *Mol Cell,* 11, 275-82.

DONACHIE, W. D. (1968) Relationship between cell size and time of initiation of DNA replication. *Nature,* 219, 1077-9.

DUGGIN, I. G. & BELL, S. D. (2006) The chromosome replication machinery of the Archaeon Sulfolobus solfataricus. *J Biol Chem.*

EDGELL, D. R. & DOOLITTLE, W. F. (1997) Archaea and the origin(s) of DNA replication proteins. *Cell,* 89, 995-8.

ERZBERGER, J. P., PIRRUCCELLO, M. M. & BERGER, J. M. (2002) The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. *Embo J,* 21, 4763-73.

FELCZAK, M. M. & KAGUNI, J. M. (2004) The box VII motif of Escherichia coli DnaA protein is required for DnaA oligomerization at the E. coli replication origin. *J Biol Chem,* 279, 51156-62.

FIELDS, S. & SONG, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature,* 340, 245-6.

FLETCHER, R. J., BISHOP, B. E., LEON, R. P., SCLAFANI, R. A., OGATA, C. M. & CHEN, X. S. (2003) The structure and function of MCM from archaeal M. Thermoautotrophicum. *Nat Struct Biol,* 10, 160-7.

FLETCHER, R. J., SHEN, J., GOMEZ-LLORENTE, Y., MARTIN, C. S., CARAZO, J. M. & CHEN, X. S. (2005) Double hexamer disruption and biochemical activities of Methanobacterium thermoautotrophicum MCM. *J Biol Chem,* 280, 42405-10.

FUJIKANE, R., KOMORI, K., SHINAGAWA, H. & ISHINO, Y. (2005) Identification of a novel helicase activity unwinding branched DNAs from the hyperthermophilic archaeon, Pyrococcus furiosus. *J Biol Chem,* 280, 12351-8.

GIETZ, R. D. & WOODS, R. A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Guide to Yeast Genetics and Molecular and Cell Biology, Pt B.*

GRABOWSKI, B. & KELMAN, Z. (2001) Autophosphorylation of archaeal Cdc6 homologues is regulated by DNA. *J Bacteriol,* 183, 5459-64.

GRAINGE, I., SCAIFE, S. & WIGLEY, D. B. (2003) Biochemical analysis of components of the pre-replication complex of Archaeoglobus fulgidus. *Nucleic Acids Res,* 31, 4888-98.

GUY, C. P. & BOLT, E. L. (2005) Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands. *Nucleic Acids Res,* 33, 3678-90.

HAYASHI, I., MORIKAWA, K. & ISHINO, Y. (1999) Specific interaction between DNA polymerase II (PolD) and RadB, a Rad51/Dmc1 homolog, in Pyrococcus furiosus. *Nucleic Acids Res,* 27, 4695-702.

HEREFORD, L. M. & HARTWELL, L. H. (1974) Sequential gene function in the initiation of Saccharomyces cerevisiae DNA synthesis. *J Mol Biol,* 84, 445-61.

HJORT, K. & BERNANDER, R. (1999) Changes in cell size and DNA content in Sulfolobus cultures during dilution and temperature shift experiments. *J Bacteriol,* 181, 5669-75.

HOLLAND, L., GAUTHIER, L., BELL-ROGERS, P. & YANKULOV, K. (2002) Distinct parts of minichromosome maintenance protein 2 associate with histone H3/H4 and RNA polymerase II holoenzyme. *Eur J Biochem,* 269, 5192-202.

HUBER, H., HOHN, M. J., RACHEL, R., FUCHS, T., WIMMER, V. C. & STETTER, K. O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature,* 417, 63-7.

HUNT, J. F., WEINKAUF, S., HENRY, L., FAK, J. J., MCNICHOLAS, P., OLIVER, D. B. & DEISENHOFER, J. (2002) Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA. *Science,* 297, 2018-26.

INOUE, H., NOJIMA, H. & OKAYAMA, H. (1990) High efficiency transformation of Escherichia coli with plasmids. *Gene,* 96, 23-8.

IYER, L. M., LEIPE, D. D., KOONIN, E. V. & ARAVIND, L. (2004) Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol,* 146, 11-31.

JACOB, F. & BRENNER, S. (1963) [On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon.]. *C R Hebd Seances Acad Sci,* 256, 298-300.

JENKINSON, E. R. & CHONG, J. P. (2006) Minichromosome maintenance helicase activity is controlled by N- and C-terminal motifs and requires the ATPase domain helix-2 insert. *Proc Natl Acad Sci U S A,* 103, 7613-8.

JOHNSON, E. R. & MCKAY, D. B. (1999) Crystallographic structure of the amino terminal domain of yeast initiation factor 4A, a representative DEAD-box RNA helicase. *Rna,* 5, 1526-34.

JOHNSON, R. T. & RAO, P. N. (1970) Mammalian cell fusion: induction of premature chromosome condensation in interphase nuclei. *Nature,* 226, 717-22.

KAGUNI, J. M. (1997) Escherichia coli DnaA protein: the replication initiator. *Mol Cells,* 7, 145-57.

KARLSSON, R. & FALT, A. (1997) Experimental design for kinetic analysis of protein-protein interactions with surface plasmon resonance biosensors. *J Immunol Methods,* 200, 121-33.

KASIVISWANATHAN, R., SHIN, J. H. & KELMAN, Z. (2005) Interactions between the archaeal Cdc6 and MCM proteins modulate their biochemical properties. *Nucleic Acids Res,* 33, 4940-50.

KASIVISWANATHAN, R., SHIN, J. H. & KELMAN, Z. (2006) DNA Binding by the Methanothermobacter thermautotrophicus Cdc6 Protein Is Inhibited by the Minichromosome Maintenance Helicase. *J Bacteriol,* 188, 4577-80.

KASIVISWANATHAN, R., SHIN, J. H., MELAMUD, E. & KELMAN, Z. (2004) Biochemical characterization of the Methanothermobacter thermautotrophicus minichromosome maintenance (MCM) helicase N-terminal domains. *J Biol Chem,* 279, 28358-66.

KATO, J. (2005) Regulatory network of the initiation of chromosomal replication in Escherichia coli. *Crit Rev Biochem Mol Biol,* 40, 331-42.

KATO, J. & KATAYAMA, T. (2001) Hda, a novel DnaA-related protein, regulates the replication cycle in Escherichia coli. *Embo J,* 20, 4253-62.

KELMAN, L. M. & KELMAN, Z. (2003) Archaea: an archetype for replication initiation studies? *Mol Microbiol,* 48, 605-15.

KELMAN, Z., LEE, J. K. & HURWITZ, J. (1999) The single minichromosome maintenance protein of Methanobacterium thermoautotrophicum DeltaH contains DNA helicase activity. *Proc Natl Acad Sci U S A,* 96, 14783-8.

KIKUMA, T., OHTSU, M., UTSUGI, T., KOGA, S., OKUHARA, K., EKI, T., FUJIMORI, F. & MURAKAMI, Y. (2004) Dbp9p, a member of the DEAD box protein family, exhibits DNA helicase activity. *J Biol Chem,* 279, 20692-8.

KITAGAWA, R., OZAKI, T., MORIYA, S. & OGAWA, T. (1998) Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for Escherichia coli DnaA protein. *Genes Dev,* 12, 3032-43.

KLUNGSOYR, H. K. & SKARSTAD, K. (2004) Positive supercoiling is generated in the presence of Escherichia coli SeqA protein. *Mol Microbiol,* 54, 123-31.

KURIMOTO, K., MUTO, Y., OBAYASHI, N., TERADA, T., SHIROUZU, M., YABUKI, T., AOKI, M., SEKI, E., MATSUDA, T., KIGAWA, T., OKUMURA, H., TANAKA, A., SHIBATA, N., KASHIKAWA, M., AGATA, K. & YOKOYAMA, S. (2005) Crystal structure of the N-terminal RecA-like domain of a DEAD-box RNA helicase, the Dugesia japonica vasa-like gene B protein. *J Struct Biol,* 150, 58-68.

LABIB, K., DIFFLEY, J. F. & KEARSEY, S. E. (1999) G1-phase and B-type cyclins exclude the DNA-replication factor Mcm4 from the nucleus. *Nat Cell Biol,* 1, 415-22.

LIANG, C. & STILLMAN, B. (1997) Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in cdc6 mutants. *Genes Dev,* 11, 3375-86.

LIM, J., THOMAS, T. & CAVICCHIOLI, R. (2000) Low temperature regulated DEAD-box RNA helicase from the Antarctic archaeon, Methanococcoides burtonii. *J Mol Biol,* 297, 553-67.

LINDER, P., LASKO, P. F., ASHBURNER, M., LEROY, P., NIELSEN, P. J., NISHI, K., SCHNIER, J. & SLONIMSKI, P. P. (1989) Birth of the D-E-A-D box. *Nature,* 337, 121-2.

LOBRY, J. R. (1996) Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol,* 13, 660-5.

LOPEZ, P., PHILIPPE, H., MYLLYKALLIO, H. & FORTERRE, P. (1999) Identification of putative chromosomal origins of replication in Archaea. *Mol Microbiol,* 32, 883-6.

LORSCH, J. R. & HERSCHLAG, D. (1998) The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry,* 37, 2180-93.

LUNDGREN, M., ANDERSSON, A., CHEN, L., NILSSON, P. & BERNANDER, R. (2004) Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A,* 101, 7046-51.

MADIGAN, M. T., MARTINKO, J. M. & BROCK, T. D. (2003) *Brock, biology of microorganisms,* Upper Saddle River, N.J., Prentice Hall.

MARINSEK, N., BARRY, E. R., MAKAROVA, K. S., DIONNE, I., KOONIN, E. V. & BELL, S. D. (2006) GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep,* 7, 539-45.

MASUMOTO, H., MURAMATSU, S., KAMIMURA, Y. & ARAKI, H. (2002) S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. *Nature,* 415, 651-5.

MATSUNAGA, F., FORTERRE, P., ISHINO, Y. & MYLLYKALLIO, H. (2001) In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc Natl Acad Sci U S A,* 98, 11152-7.

MCGUFFIN, L. J., BRYSON, K. & JONES, D. T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics,* 16, 404-5.

MENDEZ, J., ZOU-YANG, X. H., KIM, S. Y., HIDAKA, M., TANSEY, W. P. & STILLMAN, B. (2002) Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. *Mol Cell,* 9, 481-91.

MORIGEN, BOYE, E., SKARSTAD, K. & LOBNER-OLESEN, A. (2001) Regulation of chromosomal replication by DnaA protein availability in Escherichia coli: effects of the datA region. *Biochim Biophys Acta,* 1521, 73-80.

MURRAY, A. W. & HUNT, T. (1993) *The cell cycle : an introduction,* New York, Freeman.

MYLLYKALLIO, H., LOPEZ, P., LOPEZ-GARCIA, P., HEILIG, R., SAURIN, W., ZIVANOVIC, Y., PHILIPPE, H. & FORTERRE, P. (2000) Bacterial mode

of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science,* 288, 2212-5.
NISHITANI, H. & LYGEROU, Z. (2002) Control of DNA replication licensing in a cell cycle. *Genes Cells,* 7, 523-34.
NISHITANI, H., TARAVIRAS, S., LYGEROU, Z. & NISHIMOTO, T. (2001) The human licensing factor for DNA replication Cdt1 accumulates in G1 and is destabilized after initiation of S-phase. *J Biol Chem,* 276, 44905-11.
NURSE, P. (1990) Universal control mechanism regulating onset of M-phase. *Nature,* 344, 503-8.
NURSE, P. (1994) Ordering S phase and M phase in the cell cycle. *Cell,* 79, 547-50.
PAZIN, M. J. & KADONAGA, J. T. (1997) SWI2/SNF2 and related proteins: ATP-driven motors that disrupt protein-DNA interactions? *Cell,* 88, 737-40.
PRADHAN, A., CHAUHAN, V. S. & TUTEJA, R. (2005) A novel 'DEAD-box' DNA helicase from Plasmodium falciparum is homologous to p68. *Mol Biochem Parasitol,* 140, 55-60.
PRAKASH, R., KREJCI, L., VAN KOMEN, S., ANKE SCHURER, K., KRAMER, W. & SUNG, P. (2005) Saccharomyces cerevisiae MPH1 gene, required for homologous recombination-mediated mutation avoidance, encodes a 3' to 5' DNA helicase. *J Biol Chem,* 280, 7854-60.
RAO, P. N. & JOHNSON, R. T. (1970) Mammalian cell fusion: studies on the regulation of DNA synthesis and mitosis. *Nature,* 225, 159-64.
RAYCHAUDHURI, S., BYERS, R., UPTON, T. & EISENBERG, S. (1997) Functional analysis of a replication origin from Saccharomyces cerevisiae: identification of a new replication enhancer. *Nucleic Acids Res,* 25, 5057-64.
RIBER, L. & LOBNER-OLESEN, A. (2005) Coordinated replication and sequestration of oriC and dnaA are required for maintaining controlled once-per-cell-cycle initiation in Escherichia coli. *J Bacteriol,* 187, 5605-13.
ROBERTS, J. & PARK, J. S. (2004) Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination. *Curr Opin Microbiol,* 7, 120-5.
ROBINSON, N. P. & BELL, S. D. (2005) Origins of DNA replication in the three domains of life. *Febs J,* 272, 3757-66.
ROBINSON, N. P., DIONNE, I., LUNDGREN, M., MARSH, V. L., BERNANDER, R. & BELL, S. D. (2004) Identification of two origins of replication in the single chromosome of the archaeon Sulfolobus solfataricus. *Cell,* 116, 25-38.
ROCAK, S. & LINDER, P. (2004) DEAD-box proteins: the driving forces behind RNA metabolism. *Nat Rev Mol Cell Biol,* 5, 232-41.
SAMBROOK, J. & RUSSELL, D. W. (2001) *Molecular cloning : a laboratory manual,* Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.
SANCHEZ, M., CALZADA, A. & BUENO, A. (1999) The Cdc6 protein is ubiquitinated in vivo for proteolysis in Saccharomyces cerevisiae. *J Biol Chem,* 274, 9092-7.
SAXENA, S. & DUTTA, A. (2005) Geminin-Cdt1 balance is critical for genetic stability. *Mutat Res,* 569, 111-21.
SCHAECHTER, M., MAALOE, O. & KJELDGAARD, N. O. (1958) Dependency on medium and temperature of cell size and chemical composition during balanced grown of Salmonella typhimurium. *J Gen Microbiol,* 19, 592-606.
SHARPLES, G. J., INGLESTON, S. M. & LLOYD, R. G. (1999) Holliday junction processing in bacteria: insights from the evolutionary conservation of RuvABC, RecG, and RusA. *J Bacteriol,* 181, 5543-50.

SHECHTER, D. F., YING, C. Y. & GAUTIER, J. (2000) The intrinsic DNA helicase activity of Methanobacterium thermoautotrophicum delta H minichromosome maintenance protein. *J Biol Chem,* 275, 15049-59.

SHIN, J. H., GRABOWSKI, B., KASIVISWANATHAN, R., BELL, S. D. & KELMAN, Z. (2003) Regulation of minichromosome maintenance helicase activity by Cdc6. *J Biol Chem,* 278, 38059-67.

SIMMONS, L. A., FELCZAK, M. & KAGUNI, J. M. (2003) DnaA Protein of Escherichia coli: oligomerization at the E. coli chromosomal origin is required for initiation and involves specific N-terminal amino acids. *Mol Microbiol,* 49, 849-58.

SINGLETON, M. R., MORALES, R., GRAINGE, I., COOK, N., ISUPOV, M. N. & WIGLEY, D. B. (2004) Conformational changes induced by nucleotide binding in Cdc6/ORC from Aeropyrum pernix. *J Mol Biol,* 343, 547-57.

SINGLETON, M. R. & WIGLEY, D. B. (2002) Modularity and specialization in superfamily 1 and 2 helicases. *J Bacteriol,* 184, 1819-26.

SIVARAMAN, J., SAUVE, V., MATTE, A. & CYGLER, M. (2002) Crystal structure of Escherichia coli glucose-1-phosphate thymidylyltransferase (RffH) complexed with dTTP and Mg2+. *J Biol Chem,* 277, 44214-9.

SMITH, D. R., DOUCETTE-STAMM, L. A., DELOUGHERY, C., LEE, H., DUBOIS, J., ALDREDGE, T., BASHIRZADEH, R., BLAKELY, D., COOK, R., GILBERT, K., HARRISON, D., HOANG, L., KEAGLE, P., LUMM, W., POTHIER, B., QIU, D., SPADAFORA, R., VICAIRE, R., WANG, Y., WIERZBOWSKI, J., GIBSON, R., JIWANI, N., CARUSO, A., BUSH, D., REEVE, J. N. & ET AL. (1997) Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics. *J Bacteriol,* 179, 7135-55.

SNYDER, M., HE, W. & ZHANG, J. J. (2005) The DNA replication factor MCM5 is essential for Stat1-mediated transcriptional activation. *Proc Natl Acad Sci U S A,* 102, 14539-44.

SPECK, C. & MESSER, W. (2001) Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *Embo J,* 20, 1469-76.

SPELLMAN, P. T., SHERLOCK, G., ZHANG, M. Q., IYER, V. R., ANDERS, K., EISEN, M. B., BROWN, P. O., BOTSTEIN, D. & FUTCHER, B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell,* 9, 3273-97.

STANLEY, L. K., SEIDEL, R., VAN DER SCHEER, C., DEKKER, N. H., SZCZELKUN, M. D. & DEKKER, C. (2006) When a helicase is not a helicase: dsDNA tracking by the motor protein EcoR124I. *Embo J,* 25, 2230-9.

STILLMAN, B. (2005) Origin recognition and the chromosome cycle. *FEBS Lett,* 579, 877-84.

STORY, R. M., LI, H. & ABELSON, J. N. (2001) Crystal structure of a DEAD box protein from the hyperthermophile Methanococcus jannaschii. *Proc Natl Acad Sci U S A,* 98, 1465-70.

SUBRAMANYA, H. S., BIRD, L. E., BRANNIGAN, J. A. & WIGLEY, D. B. (1996) Crystal structure of a DExx box DNA helicase. *Nature,* 384, 379-83.

TANNER, N. K., CORDIN, O., BANROQUES, J., DOERE, M. & LINDER, P. (2003) The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol Cell,* 11, 127-38.

THEIS, K., SKORVAGA, M., MACHIUS, M., NAKAGAWA, N., VAN HOUTEN, B. & KISKER, C. (2000) The nucleotide excision repair protein UvrB, a helicase-like enzyme with a catch. *Mutat Res,* 460, 277-300.

THOMA, N. H., CZYZEWSKI, B. K., ALEXEEV, A. A., MAZIN, A. V., KOWALCZYKOWSKI, S. C. & PAVLETICH, N. P. (2005) Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. *Nat Struct Mol Biol,* 12, 350-6.

TSU, C. A., KOSSEN, K. & UHLENBECK, O. C. (2001) The Escherichia coli DEAD protein DbpA recognizes a small RNA hairpin in 23S rRNA. *Rna,* 7, 702-9.

TUTEJA, N. & TUTEJA, R. (2004) Unraveling DNA helicases. Motif, structure, mechanism and function. *Eur J Biochem,* 271, 1849-63.

WHITMAN, W. B., PFEIFER, F., BLUM, P. & KLEIN, A. (1999) What archaea have to tell biologists. *Genetics,* 152, 1245-8.

WOESE, C. R. & FOX, G. E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A,* 74, 5088-90.

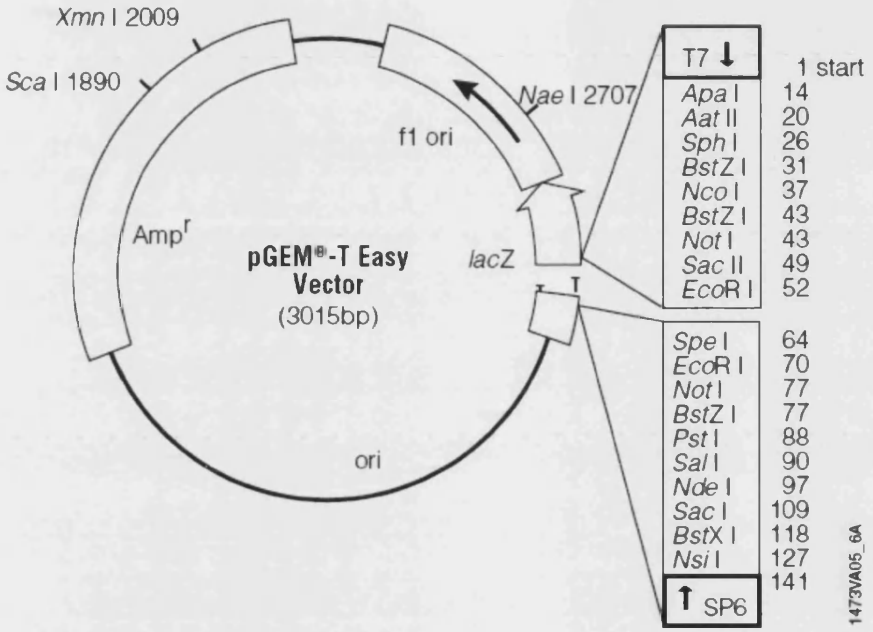
WOESE, C. R., KANDLER, O. & WHEELIS, M. L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A,* 87, 4576-9.

ZHANG, R. & ZHANG, C. T. (2003) Multiple replication origins of the archaeon Halobacterium species NRC-1. *Biochem Biophys Res Commun,* 302, 728-34.

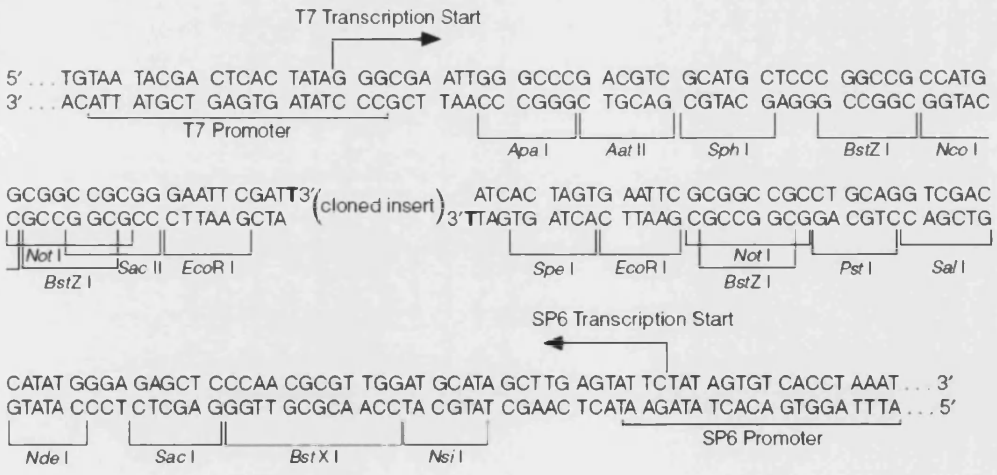
10 Appendix

10.1 Vector Maps

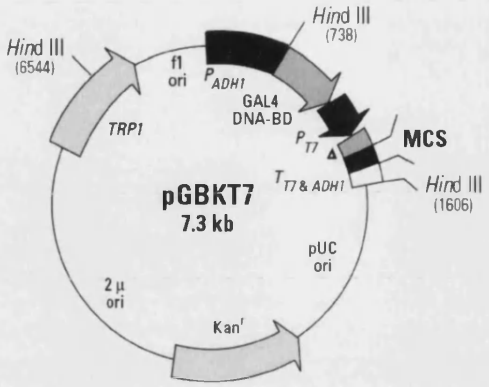
10.1.1 pGEM-T Easy



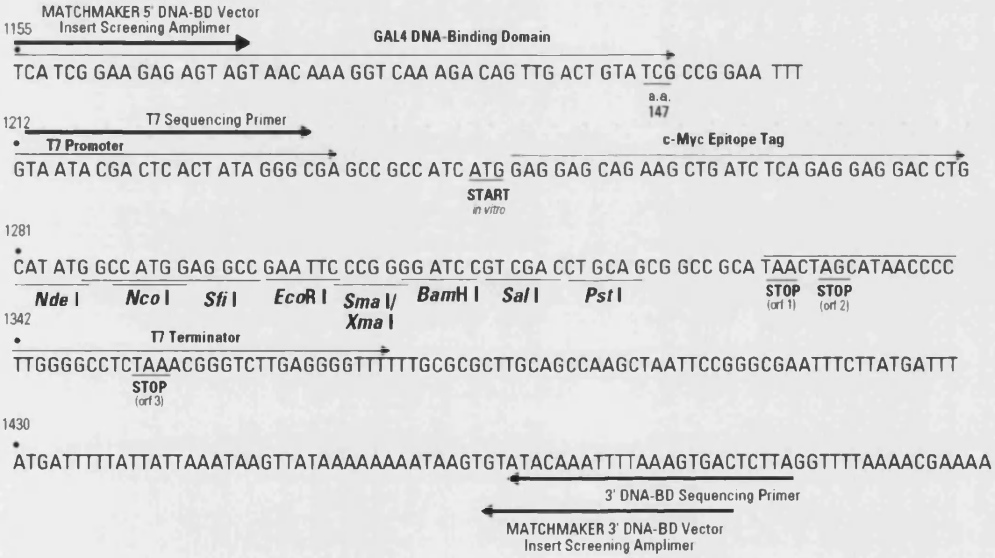
pGEM®-T Easy Vector circle map and sequence reference points.



10.1.2 pGBK-T7



▲ c-Myc epitope tag

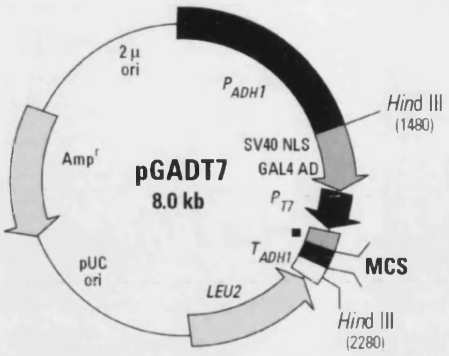


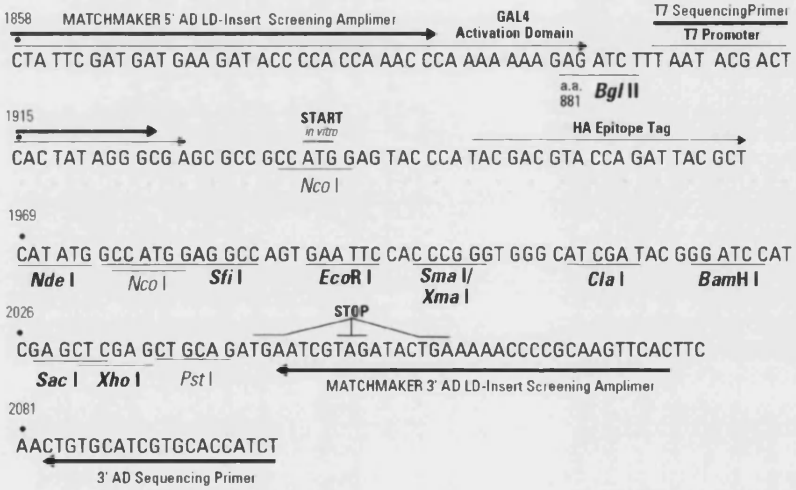
Restriction Map and Multiple Cloning Site (MCS) of pGBKT7. Unique restriction sites are in bold.

Description:

The pGBKT7 vector expresses proteins fused to amino acids 1–147 of the GAL4 DNA binding domain (DNA-BD). In yeast, fusion proteins are expressed at high levels from the constitutive *ADH1* promoter (P_{ADH1}); transcription is terminated by the T7 and *ADH1* transcription termination signals ($T_{T7\ \&\ ADH1}$). pGBKT7 also contains the T7 promoter, a c-Myc epitope tag, and a MCS. pGBKT7 replicates autonomously in both *E. coli* and *S. cerevisiae* from the pUC and 2 μ ori, respectively. The vector carries the Kan^r for selection in *E. coli* and the *TRP1* nutritional marker for selection in yeast. Yeast strains containing pGBKT7 exhibit a higher transformation efficiency than strains carrying other DNA-BD domain vectors (1).

10.1.3 pGAD-T7



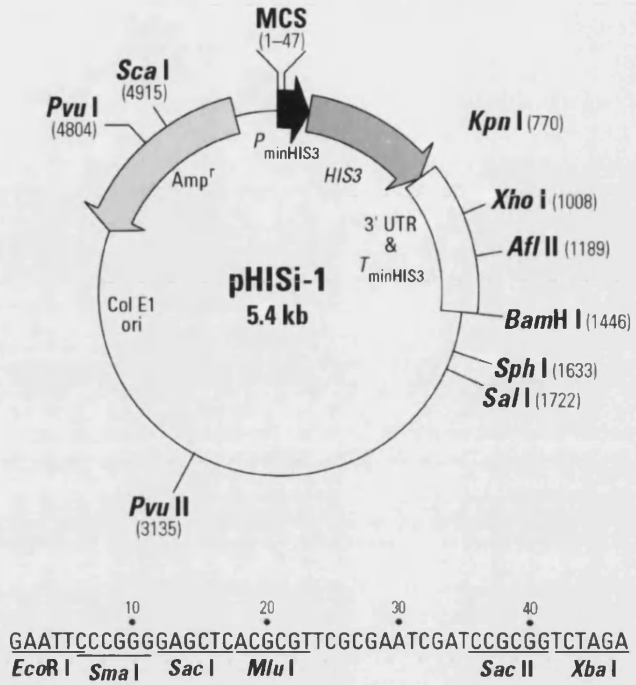


Restriction Map and Multiple Cloning Site (MCS) of pGADT7. Unique restriction sites are in bold.

Description:
The pGADT7 vector expresses proteins fused to amino acids 768–881 of the GAL4 activation domain (AD). In yeast, fusion proteins are expressed at high levels from the constitutive *ADH1* promoter (P_{ADH1}); transcription is terminated at the *ADH1* transcription termination signal (T_{ADH1}). The fusion protein is targeted to the yeast nucleus by the SV40 nuclear localization sequences that have been added to the activation domain sequence (1). pGADT7 also contains the T7 promoter, an HA epitope tag, and a MCS. pGADT7 replicates autonomously in both *E. coli* and *S. cerevisiae* from the pUC and 2 µ ori, respectively. The vector carries Amp^r for selection in *E. coli* and the *LEU2* nutritional marker for selection in yeast.

Use:
pGADT7 is the AD Vector included with MATCHMAKER Two-Hybrid System 3. The MCS of pGADT7 has unique restriction sites in frame with the 3'-end of the GAL4 AD for constructing a fusion protein with either a protein of interest or a fusion protein library. The bait protein is also expressed

10.1.4 pHisi-1



Restriction Map and Multiple Cloning Site (MCS) of pHISi-1. Unique restriction sites are in bold. The *Xho* I or *Afl* II sites can be used to stably integrate pHISi into the yeast host's genome. The *Kpn* I site cannot be used for integration because that region of the *HIS3* gene is deleted in the nonfunctional *HIS3* gene in YM4271.

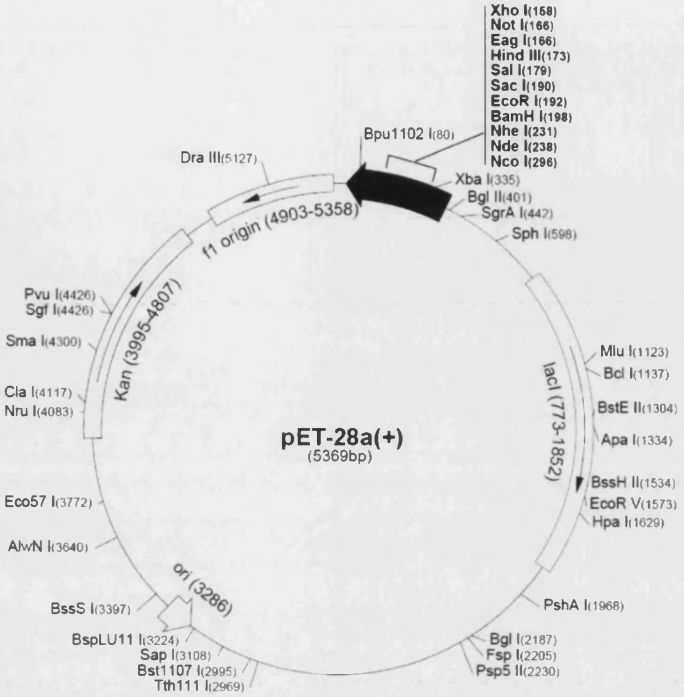
Description:

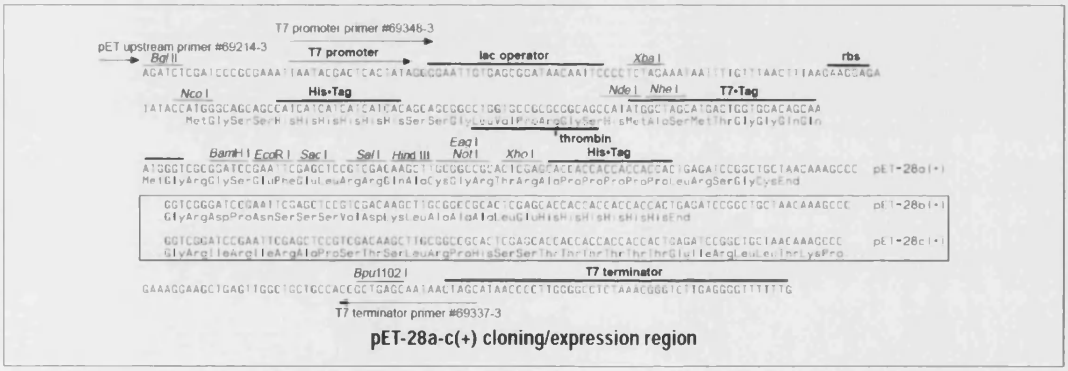
pHISi-1 is a yeast integration and reporter vector for use with the MATCHMAKER One-Hybrid System. pHISi-1 contains the yeast *HIS3* gene downstream of the MCS. P_{minHIS} is the minimal promoter of the *HIS3* locus. *Cis*-acting sequences of interest (i.e., target elements) can be cloned into the MCS. The yeast *HIS3* gene allows integration into the nonfunctional *his* locus of YM4271. Without activation by a target element, constitutive *HIS3* expression from P_{minHIS} is very low in yeast, but allows enough growth to select for integration when constucting *HIS3* reporter strains. The *HIS3* expression from pHISi-1 is generally lower than that from the pHISi Vector due to a weak UAS (upstream activating sequence) in the flanking vector sequence. During library screening, the leaky expression of *HIS3* is controlled by adding 3-AT to the medium. pHISi-1 contains the bacterial Col E1 origin (ori) and the ampicillin-resistance gene (Amp^r) for propagation and selection in *E. coli*. To construct pHISi-1, the *Eco*R I and *Bam*H I fragment, which contains the *HIS3* reporter gene, was transferred from pHISi to pBR322 at the *Eco*R I and *Bam*H I sites. pHISi-1 can be used as an alternative to pHISi and in experiments using dual reporter genes. pHISi-1 cannot replicate autonomously in yeast.

10.1.5 pET-28a

pET-28a(+) sequence landmarks	
T7 promoter	370-386
T7 transcription start	369
His• Tag coding sequence	270-287
T7• Tag coding sequence	207-239
Multiple cloning sites (*Bam*H I - *Xho* I)	158-203
His• Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	773-1852
pBR322 origin	3286
Kan coding sequence	3995-4807
f1 origin	4903-5358

The maps for pET-28b(+) and pET-28c(+) are the same as pET-28a(+) (shown) with the following exceptions: pET-28b(+) is a 5368bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-28c(+) is a 5367bp plasmid; subtract 2bp from each site beyond *Bam*H I at 198.

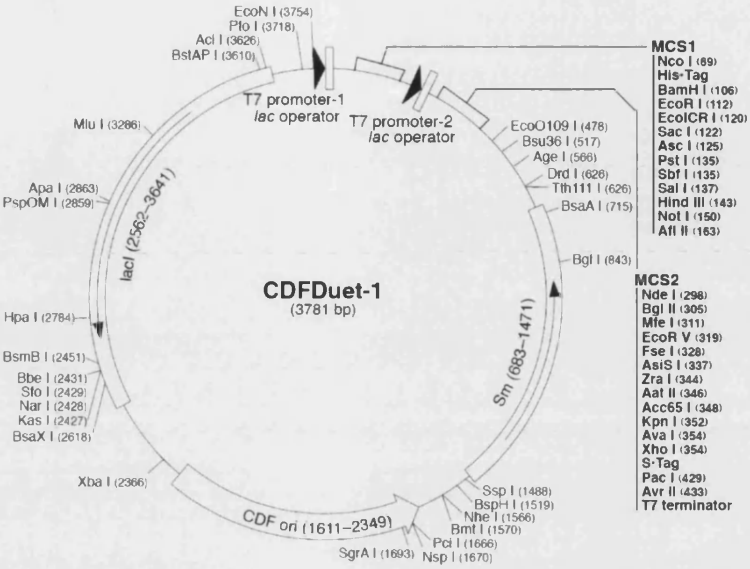


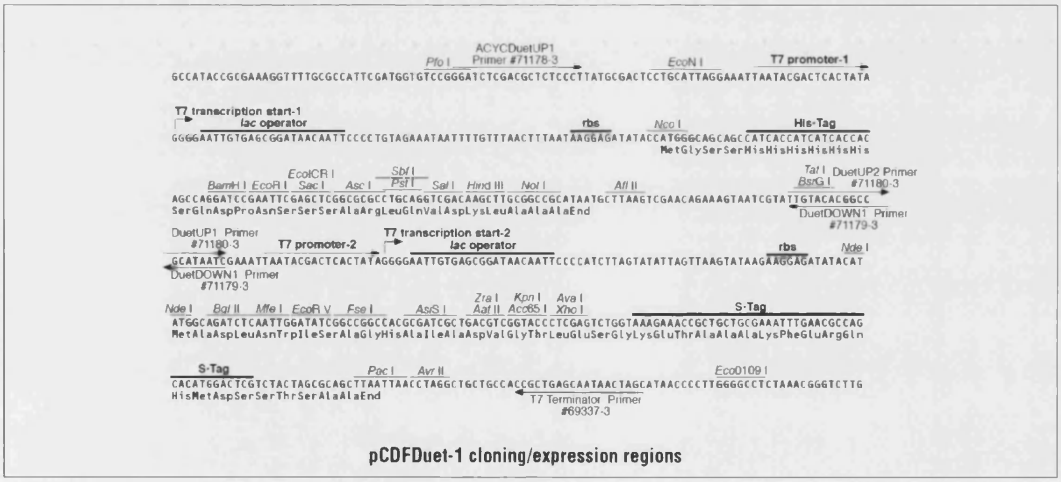


pET-28a-c(+) cloning/expression region

10.1.6 pCDF-1 Duet

Multiple cloning sites-2	
(*Nde* I–*Avr* II)	297–438
S•Tag™ coding sequence	366–410
T7 terminator	462–509
aadA (SmR) coding sequence	683–1471
CDF origin	1611–2349
lacI coding sequence	2562–3641





pCDFDuet-1 cloning/expression regions